Remote Sensing Applications: Society and Environment 27 (2022) 100820



Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase





Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2

Akiko Elders [a,b,*], Mark L. Carroll [c], Christopher S.R. Neigh [a], Anthony Louis D'Agostino [d], Christopher Ksoll [e], Margaret R. Wooten [a,f], Molly E. Brown [g]

[a] Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA
[b] Morgan State University, Baltimore, MD, 21251, USA
[c] Data Science Group, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA
[d] Mathematica, 505 14th Street, Oakland, CA, 94612, USA
[e] Department of Economics, Université de Sherbrooke, 2500, boulevard de l'Université, pavillon K1 Sherbrooke, QC J1K 2R1, Canada
[f] Science Systems and Applications Inc, Lanham, MD, 20706, USA
[g] Department of Geographical Sciences, University of Maryland College Park, MD, 20742, USA

ARTICLE INFO

Keywords:
Land cover classification
Africa
Smallholder fields
Sentinel-2
Multi-day imagery
Machine learning
Agriculture
Yield

ABSTRACT

Remote Sensing affords the opportunity to monitor and evaluate data scarce regions where field collection efforts are costly. A particular challenge is monitoring and evaluation in regions with smallholder agricultural systems (~1 ha) that are often subsistence focused, vulnerable to food insecurity and data scarce. Using multi-day moderate resolution Sentinel-2 and Random Forest models, this study shows that crop type and rice yields in Burkina Faso can be predicted with greater than ~80% accuracy in the rainy season. Model optimization using varying spectral and vegetation index inputs can increase crop type and yield prediction accuracy in the dry season where denser cultivation is a challenge for the 10–20 m resolution of Sentinel-2. However, there is a trade-off between opting for very high-resolution imagery (<2 m) or the number of bands offered by Sentinel-2 as the bands that occupy and vegetation indices that utilize the red through NIR ranges were most important across all models. In addition, model type, linear Regression or nonlinear Random Forest, matters little when estimating yield in these landscapes, unless Harmonic regression is utilized for the linear model. This study also showed that a model trained with high quality 2019 dry season crop cut data can predict the subsequent dry season's interannual crop type with overall accuracy as high as 60%, comparable to crop type models trained with 2020 survey data and used to estimate crop type in the concurrent season, as the survey collection. This indicates some utility in leveraging the calibrated Random Forest models to make skillful predictions of interannual crop type and ultimately food availability of nearby communities for years with no training data. Given increasing global food prices and restricted commodity trade, understanding local agricultural productivity using affordable and timely remote sensing-based methods is essential for ensuring appropriate humanitarian interventions.

* Corresponding author. Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.
E-mail addresses: akiko.elders@nasa.gov (A. Elders), mark.carroll@nasa.gov (M.L. Carroll), christopher.s.neigh@nasa.gov (C.S.R. Neigh), adagostino@mathematica-mpr.com (A.L. D'Agostino), christopher.ksoll@usherbrooke.ca (C. Ksoll), margaret.wooten@nasa.gov (M.R. Wooten), mbrown52@umd.edu (M.E. Brown).

https://doi.org/10.1016/j.rsase.2022.100820
Received 18 May 2022; Received in revised form 28 July 2022; Accepted 29 July 2022
Available online 2 August 2022
2352-9385/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Background

Earth observing satellites provide a unique opportunity to monitor and evaluate crop density, composition and yield with spatial and temporal resolutions that facilitate global monitoring (Chang et al., 2007; Doraiswamy, 2004; McCarty et al., 2017; McNally et al., 2015; Neigh et al., 2018). Remotely sensed data has been extensively used to provide advanced warning of drought and other agroclimatic shocks (Verdin et al., 2005; Brown et al., 2007) that reduce food availability, affecting the food security of between 720 and 811 million people globally (Brown et al., 2007; FAO, 2021). Earth observing systems are especially useful for remote data collection in regions where agricultural data is limited, farmer self-reporting results in bias, and field collection efforts are costly (Burke and Lobell, 2017; Carletto et al., 2017; Lobell et al., 2020).

A challenge however, is monitoring productivity where smallholder agriculture comprises a sizable proportion of agricultural activities, such as in Africa, Asia and parts of Europe (Lesiv et al., 2019; Lowder, 2016; Fritz et al., 2011). In these cases, satellite sensors with the longest archives such as Landsat (50 years, at 30 m resolution) and MODIS (20 years, at 250 m resolution) have spatial resolutions that are insufficient in capturing the nuances of these small fields. This is important, for example, in African nations where 75 percent of the total fields may be less than 0.5ha and are often dominated by mixed land uses (Burke and Lobell, 2017; Carletto et al., 2015; Debats et al., 2016). In regions with high variability of precipitation, smallholder farmers who rely heavily on rain fed agriculture are particularly vulnerable to food insecurity (Niles and Brown, 2017; Rockström et al., 2003).

Satellite systems such as Sentinel-2 (S2) A and B, with 10 m resolution, weekly revisit frequency and publicly accessible data through cloud computing platforms, such as Google Earth Engine, offer a solution for the evaluation of regions with smallholder agriculture (Burke and Lobell, 2017; Drusch et al., 2012). The use of high-resolution (2–8 m resolution) satellites to evaluate crop composition and yield has been examined. Overall, accuracy is improved as the spatial resolution increases for both crop classification (Rao et al., 2021; Turker and Ozdarici, 2011) and yield (Burke and Lobell, 2017). Higher resolution imagery is particularly beneficial in regions with very small field sizes where yield predictions are nearly as skillful as measurements obtained through surveys (Burke and Lobell, 2017).

Empirical methods for predicting crop composition use measured parameters such as spectral signature, derived indices, or texture, often acquired from only one or a select number of acquisition dates, in combination with machine learning models such as Random Forests or Support Vector Machines (SVM) (Aguilar et al., 2018; Chellasamy et al., 2014; Fan et al., 2021a, 2021b; Gerstmann et al., 2016; Kim and Yeom, 2014; Palchowdhuri et al., 2018; Ursani et al., 2012; Zafari et al., 2017; Zurita-Milla et al., 2017). However, a number of studies have shown the utility of incorporating temporal information from multiple dates, optimally 5, for crop classification to capture field spectral and textural variations. To include temporal information, numerous studies have employed temporal windows, periods of time within the growing season that target specific periods of crop phenology (Conrad et al., 2014; Debats et al., 2016; Waldhoff et al., 2017) and time series of derived spectral properties (Inglada et al., 2015; Valero et al., 2015, 2016; Wardlow and Egbert, 2008). However, the need to determine an optimal temporal window can be eliminated by including multiple dates throughout the growing season leading up to months with peak crop development, where peak and late season imagery are more valuable for model classification skill than incorporating early season imagery (Rao et al., 2021; Vuolo et al., 2018), since the difference between natural vegetation and crops are in part their length of growing period and senescence date. In regions with small field agriculture, changes in texture become more important than changes in spectral characteristics, and increased temporal information is more advantageous than additional spectral bands (Debats et al., 2016).

Similarly, observed yields are related to remotely sensed parameters such as reflectance and derived vegetation indices in effort to understand production at both field and regional scales (Weiss et al., 2020). Commonly used approaches relate derived metrics such as vegetation indices (Burke and Lobell, 2017; Lobell et al., 2020) to observed yields for example, employing interpolated time series (Inglada et al., 2015; Quarmby et al., 1993; Son et al., 2014) and seasonal growth profiles (Labus et al., 2002) in combination with regression, machine and deep learning methods. Machine and deep learning are ideal for understanding nonlinear quantities such as crop yield (Palanivel and Surianarayanan 2019). Commonly utilized machine learning models for estimating yield include, random forests, linear regression, support vector machine, and gradient boosting trees such as XGBoost, where neural networks have been employed the most in previous studies (dela Torre et al., 2021; Klompenburg, Kassahun and Catal 2020; Ma et al., 2019; Palanivel and Surianarayanan 2019). However, no one model outperforms the others in terms of its prediction skill; a variety of models are often tested to better understand yield estimation accuracies where machine learning models are selected based on factors such as ease of computation, methodology and data structure (Klompenburg, Kassahun and Catal 2020; Maxwell et al., 2018). In addition, data quantity often influences model selection for example; Random Forest, an ensemble of decision trees is less affected by sample size than a single tree or Artificial Neural Networks (Maxwell et al., 2018). Of particular interest at both field and regional spatial scales is harvest predictability and accurate measures of production (Lobell et al., 2020; Weiss et al., 2020). Most previous work has focused on yield estimation in regions with homogenous land cover types and large field sizes (Weiss et al., 2020). Heterogeneous regions with a high degree of within-field variability and mixed vegetation types such as agroforestry present a challenge, which can limit yield estimation skill (Inglada et al., 2015; Schut et al., 2018; Weiss et al., 2020). However, even with finer resolution cloud-free imagery, accurate in situ yield measurements are needed to transform relative production estimates into absolute estimates per hectare, particularly in tropical regions (Weiss et al., 2020). In situ yield measurement error affects accurate yield estimation where harvest measurements reliant on farmer reported yields are biased toward higher yields with little correlation to more precise measures of harvested yield such as full plot and crop cut based estimates (Lobell et al., 2020). Though crop cuts harvest estimates capture only a quarter of the field variability in heterogeneous agricultural systems, models built with crop cut harvest estimates better correlate to full plot estimates, the preferred in situ measurement, than self-reported yields and are a reasonable alternative when full plot estimates are not feasible (Lobell et al., 2020).

Here we focus on using satellite data to monitor agricultural variability in a smallholder system in West Africa, a region with particularly poor agricultural statistics. Food security and reliance on subsistence farming in the region remains a major socio-economic problem, with 25% of children under five being stunted and 40% of the population in Burkina Faso suffering from extreme poverty (WFP, 2022). Most people depend on the current season's agricultural production for their livelihoods, leaving them with significant climate vulnerability. Vulnerability across Burkina Faso to climate shocks is highly variable across the country, with children in some regions with highly productive agricultural regions showing malnutrition while other areas have children who are less affected (Grace et al., 2016, 2017). Household food security is shaped by the pathways through which households gain access to food. Turner et al. (2021) surveyed households in the region and report low levels of self-provisioning of grain (averaging around 50% of need) with different seasonal patterns of grain purchases based on cash sources, which are affected by international commodity prices as well as local production (Brown and Kshirsagar 2015). Burkina Faso has also experienced internal conflict from armed Islamist militant groups, displacing 1.5 million people who remain extremely food insecure and who rely upon food purchased in markets (Human Right Watch. World Report, 2022). In this context, knowing the extent of annual cropped area and crop productivity is critical to understanding the food available in the country. These statistics have not typically been available in the region, therefore new research is needed.

Previous works have examined the utility of temporal variance in datasets both within and outside African agricultural systems (Abubakar et al., 2020; Aguilar et al., 2018; Vuolo et al., 2018). Calibrated models have been utilized for other types of land cover mapping including forest cover, forest change and water detection (Carroll et al., 2017). While Wang et al. (2019) examined the skillfulness of crop type mapping without field labels in the United States, this has not been done in a smallholder setting.

This study focuses on crop type and yield prediction skill of Random Forest models using multi-temporal moderate-resolution Sentinel-2 imagery in heterogeneous smallholder agricultural landscapes. The aim of this study is to understand the composition of cropped area and crop productivity, which are statistics historically unavailable in this region, and particularly important for food security. In addition, this study compares nonlinear and linear models for estimating yield evaluating model selection for yield, with similar methodologies, in heterogeneous smallholder landscapes. A unique contribution of this study is the application of a calibrated model to assess interannual crop type predictability in an uncalibrated approach in these landscapes. We explore whether an in situ calibrated Random Forest model can predict crop type for years without field data in a region, where data is often limited, and skillful crop type mapping without field labels would be of significant value.

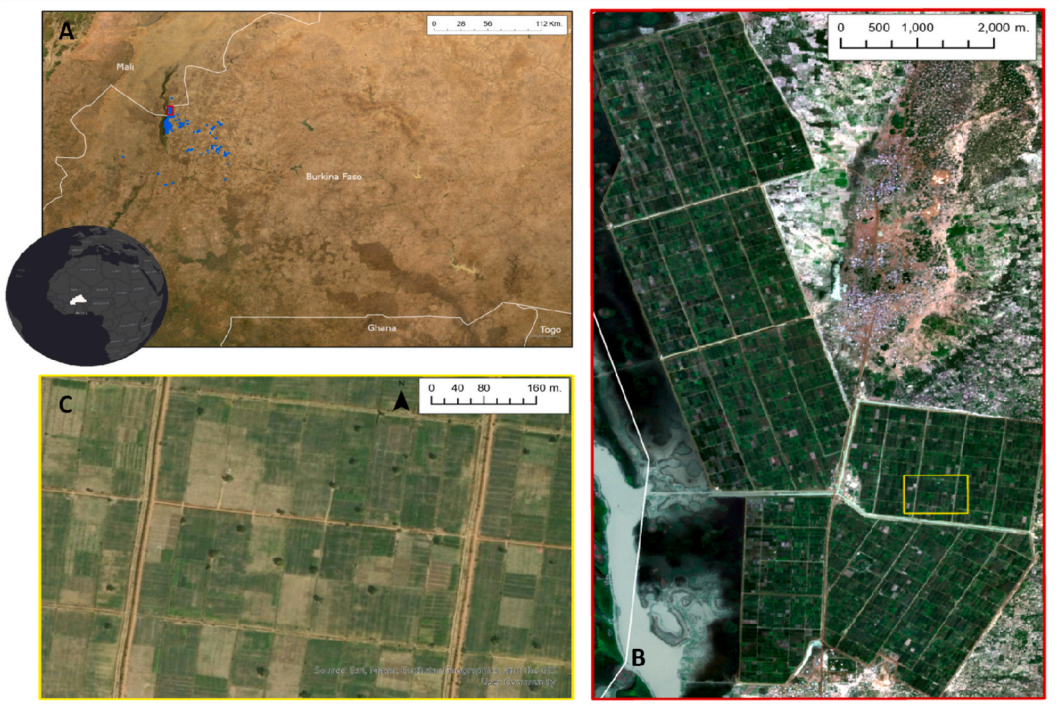


Fig. 1. Burkina Faso's location in West Africa and the study region, Di perimeter, location in Burkina Faso shown in red, overlaid with crop cut data points in blue (a). Sentinel-2 imagery, acquired for September 12, 2019, showing the Di perimeter (b) and an enlarged image showing cultivated plots and landscape heterogeneity within the perimeter (c) (High Resolution Maxar imagery obtained using ESRI's Basemap). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2. Data & methodology

2.1. Study area and crop cut In situ data: Di, Burkina Faso

The primary study region encompasses ~2246 ha of irrigated cropland located adjacent to the Burkina Faso – Mali border near the town of Di in Burkina Faso (Fig. 1b, Ksoll et al., 2021). Burkina Faso is a country located in the Sahel, dominated by high average temperature and low average rainfall, where climate variability can lead to substantial impacts on agriculture, which makes up 1/3 of the country's economy and employs 80 percent of the population either directly or indirectly (USAID Burkina Faso, 2015). The Millennium Challenge Corporation (MCC)'s Agricultural Development Project (ADP) constructed irrigation infrastructure, which was completed by mid-2014 (Fig. 1, left). It was intended to enable complementary irrigation for rainy season agriculture (May–November) and, more importantly, to enable farmers to cultivate a second season crop during the dry period (November–April). The land within the irrigated area is called the Di perimeter, and can be divided into two zones, 1) land suitable for rice cultivation in the areas adjacent to the Sourou River, and 2) slightly higher-lying land suitable for a variety of crops including maize in the rainy season and vegetables such as onions and tomatoes in the dry season. Within each zone, land was leveled and has similar access to water; either flooded for rice cultivation or distributed via small canals dug by the farmers. Soils are nutrient poor and require application of organic or inorganic fertilizer (Ksoll et al., 2019).

Fields allocated to farmers are generally less than 2 ha with the smallest field sizes in the sample are about 0.5 ha (Fig. 1c). Prior to the construction, the dominant crops were sorghum as well as maize and rice, the ADP anticipated and supported a shift away from sorghum to maize and rice in the rainy season; and to higher-value vegetable crops in the dry season (Ksoll et al., 2018). This shift was supported by training farmers in best practices for 6 priority crops: maize, rice, onions, soybeans, cowpeas, and tomatoes. Training included techniques of land preparation (double-ridging, planting, cultivating onions on high platforms), input use (fertilizer, pesticides, and herbicides) and use of irrigation (Ksoll et al., 2019).

Geo-located in situ crop type and yield samples were collected as part of an independent evaluation of these investments. More than 1500 pure stand crop cuts (5 m × 5 m) were collected for the 2018–2019 dry season (905 samples) and 2019 rainy season (612 samples) (Table 1). A stratified random sample of farmers cultivating land on the perimeter were drawn, covering 230 and 270 plots of land (out of 2029 total plots) for the dry and rainy seasons, respectively (Appendix I). Plot sizes range between 0.11 and 7.3ha, with an average plot size of 1.17ha, where only 15% of plots are larger than 2ha.

The ground data collection consisted of the ADP's 6 priority crops: maize, rice, onions, soybeans, cowpeas, and tomatoes (Ksoll et al., 2021). Data was collected through trained interviewers: after planting, during the season, and at harvest and the sampled crops were weighed (measuring dried grain) in the measurement squares. For each selected square, yield data were recorded once for crops with a single harvest and multiple times for crops with possible multiple rounds of harvesting, as in the case of crops like tomatoes. Crops like tomatoes are routinely harvested multiple times in a season, and therefore total output is the sum of harvest weights across all harvest events. Small sample issues ($N < 100$) exist for a subset of crops across either season, so our analysis focuses on maize and rice in the rainy season, and onions, tomatoes and rice in the dry season that have sufficient observations (Table 1).

The centroids of the measurement squares were used to account for measurement inaccuracies and were combined with a non-crop class (297 samples). The samples in the non-crop class consisted of features manually created by visually inspecting Very High-Resolution satellite imagery including, Maxar, using ESRI's ArcGIS software (Esri, 2009). The non-crop features were created to distinguish cultivated cropland from bare ground, water and forestry present in areas surrounding Burkina Faso's Di perimeter.

2.2. Survey based In situ data: Di and Boucle du Mouhoun

A subsequent round of field data collection was conducted in 2020 and collected survey data on crops grown during the 2019/2020 dry season, which spanned from planting in November 2019 through harvesting in March/April 2020. A total of 774 crop type samples were collected both for fields in the perimeter, as well as from plots outside of the perimeter in the Boucle du Mouhoun region and included 19 different crop types (Appendix 1). This information was collected by interviewers who visited the fields of household survey respondents and walked into the middle of the field that had been cultivated and took GPS coordinates of the cultivated field. This survey data was combined with the non-crop class to reduce error in classification due the presence of water, bare ground and agroforestry in the study region.

2.3. Satellite imagery: Sentinel-2

The cloud-based computing platform Google Earth Engine (GEE) was used to access the European Space Agency's Sentinel-2 (S2)

Table 1
The distribution of crop cut samples for each crop used to train and test the Random forest model.

Crop	Dry season (2018/2019)			Rainy season (2019)			Dry season (2019/2020)
	#	% of total	# with yield data	#	% of total	# with yield data	#
Maize	69	7.62%	39	463	75.65%	463	51
Onion	337	37.24%	278	0	N/A	0	147
Rice	162	17.90%	147	134	21.9%	128	44
Tomato	283	31.27%	190	9	1.47%	4	82
Cowpeas	54	5.97%	36	2	0.33%	2	1
Soybeans	0	NA	0	4	0.65%	3	0

level-2A surface reflectance repository (European Space Agency, 2016; Gorelick et al., 2017; Müller-Wilm, 2016). Sentinel-2 was used for both its high spatial (10 m in visible and near-infrared) and temporal resolution (5 days at the equator), which is advantageous for the small field size of the study area (Drusch et al., 2012). Multiple dates without extensive cloud cover were selected throughout the 2019 rainy (6 dates between September and November) and dry (12 dates between January and March) seasons (Table 2). The S2 cloud and cirrus cloud BitMask was used for cloud masking. Eight band S2 imagery – Red, Green, Blue, Near Infrared (NIR), 2 Short-Wave Infrared (SWIR), and 2 Red Edge (RE) – was acquired for each acquisition date during the peak to the end of the growing season.

2.4. Methodology

2.4.1. Machine learning algorithm

Random Forest classification and regression were used to estimate crop type and yield respectively, where classifiers are ideal for categorical data and regression is ideal for continuous data and extrapolations. Random Forest models use an ensemble of decision trees to determine the best outcome from a random selection of sample data and independent variables (Belgiu and Drăguţ, 2016; Breiman, 2001). In Random Forest models, bootstrap samples are created from the training data by randomly sampling with replacement. These samples are then fed to separate decision trees for training, with the result determined by majority vote from the ensemble. This method, also known as Bootstrap aggregation, improves upon a conventional decision tree's tendency to overfit by reducing variance from the original training data sample (Breiman, 2001). These models have been increasingly used in remote sensing due to their ease of implementation, overall accuracy, and speed (Belgiu and Drăguţ, 2016; Vuolo et al., 2018).

To account for spectral differences between dry and rainy seasons, separate Random Forest models were built for each. A Random Forest Classifier was first used to predict crop type. Yield was then estimated with a Random Forest Regressor and crop type masks were applied to yield results to limit the evaluation to target crops. The same training data are used for both the Random Forest Classifier and Random Forest Regressor and discussed in section 2.4.2. For this work, 100 decision trees were used and 25 percent of the data were held back for validation. Breiman (2001) demonstrated increasing the number of trees, converges model errors (to a limit) as the size of the forest grows. Additionally, fine-tuning the number of trees ensures the size of the forest is appropriate for the quantity of training data. We tested different numbers of trees, and found 100 decision trees optimized accuracy (Appendix 2 Figure A1). Similarly, testing was conducted to assess differences in accuracy achieved between reserving 10%–40% of the data for model validation. In this context, the reserved data are not used to train the model and instead test the model's skill on an unseen dataset. The greatest impact on accuracy was observed in test sizes greater than 25%, while minimal differences were otherwise observed. The goal was to maximize the data used to both train and test the models while maintaining optimal accuracy. Therefore, a test size of 25% was chosen, instead of 15% (which showed a marginal 0.1% increase in accuracy compared to 25%) in effort to allot ample data for testing purposes (Appendix 2 Figure A2). Once the testing threshold was set, the test data were randomly selected and separated from the training dataset using python's Scikit-learn package (Pedregosa et al., 2011).

We selected a set of metrics to evaluate the efficacy of each Random Forest model. For the Random Forest classifier, a confusion matrix was generated from the subset of data reserved for validation. A confusion matrix compares a crop classified by the model to its actual or reference class. In this way, errors of commission and omission as well as classification skill (producer accuracy) for each class are easily identifiable. Commission error refers to the number of crops classified in the incorrect class while omission error is for those omitted from their correct class. Additionally, the out of bag (OOB) accuracy was used to measure the Random Forest model's overall classification accuracy. OOB accuracy is computed by averaging each decision tree's ability to predict an outcome using only the out of bag sample, or the subset of data that was not used to train a particular tree (Breiman, 2001).

To evaluate the skill of yield estimates, accuracy and error were evaluated based on the residuals, or the difference between the observed and forecasted yield value. The Mean Absolute Percentage Error (MAPE) is a measure of the magnitude of the forecast error, where a larger MAPE value indicates a greater difference between the forecasted value and the actual yield value (equation (1)). To measure accuracy, MAPE was subtracted from 100 percent.

$$\widehat{MAPE} = \frac{100}{N} * \Sigma \frac{|y_i - \widehat{y_i}|}{y_i} \quad (1)$$

$\widehat{y_i}$ sample forecasted yield
y_i actual yield

Table 2
Sentinel-2 acquisition dates and bands included in the models.

Season	Bands	Resolution	Imagery date	
			RF models	Regression models
Rainy (June–Nov 2019)	Red, Green, Blue, NIR, Red Edge$_{705}$, Red Edge$_{740}$,SWIR 1, SWIR2	10m	September: 12, 22, 27 October: 22, 27 November: 01	August: 13, September 12, 22, 27October: 7, 22 November: 1, 21, 26 December: 1, 4, 11, 16, 21, 26
Dry (December 2018–March 2019)			January: 5,15,20,25,30 February: 4,9,14,24 March: 1,11,26	November: 11 December: 6, 11, 16, 21, 26 January: 15, 20, 25, 30 February: 4, 14 March: 1, 16 April: 10, 15, 20

N number of samples

In addition to MAPE, the Mean Absolute Error (MAE) was used to quantify the magnitude of error. MAE is the average absolute value of the residual, where a large MAE is indicative of large forecast error. The absolute nature of MAE means forecast errors are interpreted in the same units as the observations, tons per hectare. In this context, the MAE was used to quantify the error of the yield model's average estimated yield and treats a 0.5 t/ha error for an observation of 5 t/ha the same as an error for an observation of 3 t/ha.

$$\widehat{MAE} = \frac{1}{N} * \Sigma \left| \widehat{y_i} - y_i \right| \quad (2)$$

$\widehat{y_i}$ sample forecasted yield
y_i actual yield
N number of samples

2.4.2. Model optimization and selection

A series of five Random Forest crop type and yield models were tested to identify the optimal combination of spectral information and metrics such as vegetation indices (VI) that would increase accuracy when included in a Random Forest model. The vegetation indices were derived from S2 following the methods of Lobell et al. (2020), and include Normalized Difference Vegetation Index (NDVI), Red-Edge NDVI 705 (NDVI 705), Red-Edge NDVI 740 (NDVI 740), Green Chlorophyll Vegetation Index (GCVI), and MERIS Terrestrial Chlorophyll Index (MTCI). In addition, the seasonal amplitude, median, and maximum of vegetation indices were computed for the imagery dates acquired for each season. Seasonal amplitude is defined as the difference between the maximum and minimum seasonal index value and was evaluated for each pixel to distinguish each crop type by its unique seasonal range for all vegetation indices included in this study.

Table 3 contains detailed information about each crop type and yield Random Forest model and the spectral information and metrics included in each of the five models: BAND7NDVI, VI5, BAND7VI5, MEDMAX, and MEDMAXAMP. The skill of each Random Forest model was evaluated based on changes in accuracy and error, and the method that produced a crop type or yield model with the highest accuracy and lowest errors was selected. Therefore, a unique Random Forest model was trained and applied for this evaluation based on target season or crop type rather than applying a single model. Input features were selected to ensure model comparability between linear and nonlinear methods (discussed in section 2.4.3) as well as to maximize the spectral information available from Sentinel-2. Feature importance was detected to identify optimal bands that aided in the model's selection process among the best performing models.

2.4.3. Comparing nonlinear and linear yield models

In addition to optimizing a Random Forest model by evaluating the 5 VI-based models described above, the differences between linear and Random Forest regression models were evaluated to determine advantages and limitations of linear and nonlinear methods for estimating yield.

The first linear regression approach estimates yield (y_i) as a function of plot-specific maximum and median VI values observed over the course of the full season, for a single VI. Pixels that exceed a defined cloudiness threshold of greater than 20% and a value of 0 in the cloud and cirrus bitmask were masked (Equation (3)).

$$y_i = \beta_0 + VI_i^{max} + VI_i^{median} + \varepsilon_i \quad (3)$$

The second linear regression model leverages the joint and complementary information contained across multiple VIs augmenting the single-VI regression model (Equation (4)). Since these VIs are constructed from different spectral bands, taking a multiple-VI

Table 3
Sentinel-2 bands and derived metrics included in 5 different Random Forest models: BAND7NDVI, VI5, BAND7VI5, MEDMAX and MEDMAXAMP.

Feature	Bands/Equations	Band7NDVI	VI5	BAND7VI5	MEDMAX	MEDMAXAMP
RED (R)	B4	x		x		
Green (G)	B3	x		x		
Blue (B)	B2	x		x		
Near Infrared (NIR)	B8	x		x		
Red Edge (RE)	B5 (Red Edge 1)	x		x		
	B6 (Red Edge 2)					
Short Wave Infrared (SWIR)	B11 (SWIR1)	x		x		
	B12 (SWIR 2)	x		x		
NDVI	(NIR –R)/(NIR + R)	x	x	x		
NDVI740	(NIR -RE2)/(NIR + RE2)		x	x		
NDVI705	(NIR -RE1)/(NIR + RE1)		x	x		
GCVI	(NIR/G) −1		x	x		
MTCI	(NIR-RE1)/(RE1-R)		x	x		
Median	Seasonal median computed for all vegetation indices				x	x
Maximum	Seasonal maximum computed for all vegetation indices				x	x
Amplitude	Seasonal amplitude computed for all vegetation indices					x

approach is likely to incorporate more valuable information than a method that models yields as solely a function of an individual VI.

$$y_i = \beta_0 + \sum_v \left(VI_{iv}^{max} + VI_{iv}^{median} \right) + \varepsilon_i \quad (4)$$

for all i observations of a given crop grown in a specific season.

2.4.4. Random forest model transferability and prediction skill

For this evaluation, 2020 dry season crop type was estimated by leveraging the 2019 dry season Random Forest crop type models to estimate crop type in years without training data. This method characterizes an uncalibrated approach by applying a crop type model trained on dry season 2019 in situ crop cut data to estimate 2020 crop type for the same geographic area. Survey data for dry season 2020 was used to validate skill and determine the interannual prediction skill. Random Forest model transferability is limited to dry season crop type identification as yield data was not collected for the validation dataset and survey data was collected only during the 2020 Dry season. To maintain consistency between datasets the 2020 survey data was restricted to only crop types present in the 2019 crop type model (within the Di perimeter) and included rice, tomatoes, onions, cowpeas, maize, and non-crops. Rates of accuracy and misclassification were used to determine prediction skill. Skill metrics were derived from the confusion matrices and evaluated in two ways. 1) Overall accuracy was calculated and is defined as the sum of correctly classified crops divided by the total number of crops classified (where misclassification is those incorrectly classified). 2) Producer accuracy was calculated and is the degree to which each crop was correctly classified when compared to the total number in its reference class. It is noted that crop types with small sample sizes in the 2019 and 2020 in situ datasets, such as cowpeas, fail to classify when making interannual predictions.

3. Results

3.1. Estimating crop composition

Crop composition was estimated for both the rainy and dry season for all crops of interest. Random Forest crop type models for both the rainy and dry season were trained with the spectral signature and metrics of the crops during months that occur during the peak of

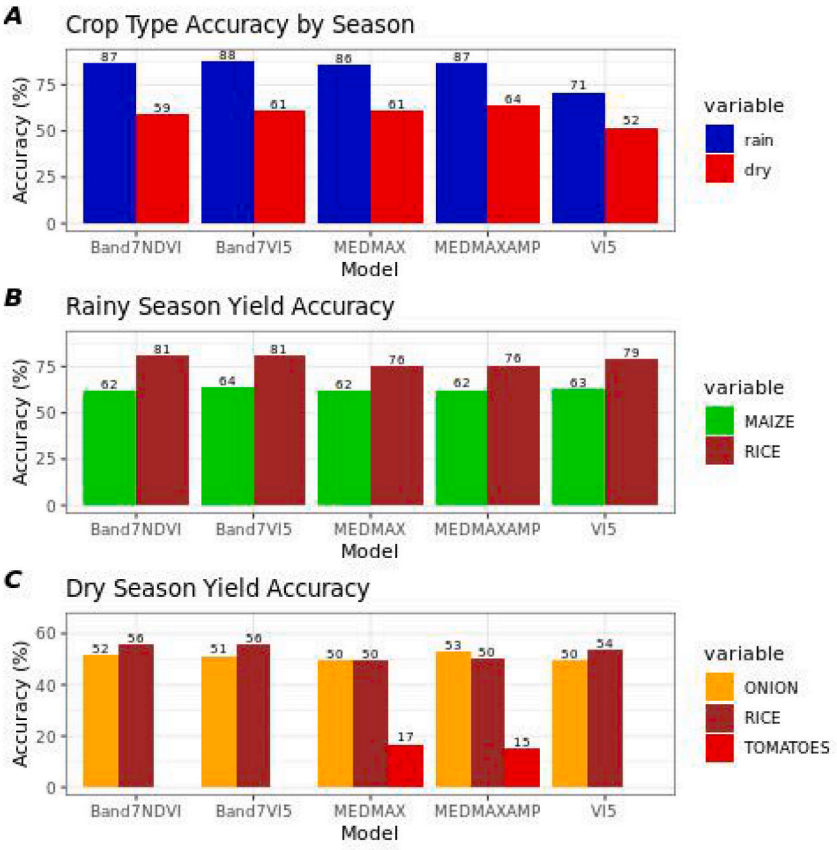


Fig. 2. The difference in model skill by season and crop type. **A** model skill when estimating crop type using Random Forest Classification by season, rainy (blue) dry (red). **B** model skill when estimating rainy season Maize (green) and rice (brown) yield. **C** model skill when estimating dry season onion (orange), rice (brown) and tomato (red) yield, where three models exhibited negative accuracy when estimating dry season tomato yield (BAND7VI5, BAND7NDVI and VI5). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the growing season. As expected, the varying input feature combinations used to train the Random Forest crop type models produced a range of outcomes, dependent on the season and crop of interest (Fig. 2a). For the rainy season, crop type accuracy increased 17% percent from the least (VI5) to the most (BAND7VI5) skillful crop type model while in the dry season, accuracy increased 12% (least skillful: VI5, most skillful: MEDMAXAMP). The differences in accuracy based on crop type model inputs for the rainy and dry seasons highlights the utility of applying unique models for each season rather than a "*one model fits all*" approach. Therefore, the following discussion will focus on the results of the BAND7VI5 model for rainy season crop type estimates and MEDMAXAMP for the dry season.

Overall, crop type classification skill was greater in the rainy season across all models, with an out-of-bag accuracy of 88% in the most skillful model, BAND7VI5. Two crop types, rice and maize were predominantly identified (Fig. 3a) and comprise 94.5% of the cultivated area (18.5% rice, 76% maize). Rice and maize were classified into the correct class 82% and 95% of the time, respectively, highlighting a high degree of classification accuracy, or Producer's accuracy (Fig. 3b). The remaining crops of tomatoes, cowpeas and soybeans underperform in the BAND7VI5 model, all three crop types are 100% omitted from their respective classes; confused for maize when misclassified. The inaccuracy in classifying tomatoes, cowpeas and soybeans is attributed to undersampling, where the total number of crop cut samples in the classes combined was less than ~3% of the data used to train the BAND7VI5 Random Forest model (Fig. 3b).

In the dry season, out-of-bag accuracy was 64% with the tomatoes, rice and onions predominantly classified, comprising ~88% of the cultivated area (Fig. 4a). Onions accounted for the largest proportion of cultivated area, covering approximately 41% of the area followed by tomatoes at 31%, and rice 16%. Tomatoes, onions and maize were spectrally less distinguishable and consistently misclassified with each other. A denser degree of intercropping and GPS error may lead to spectral mixing along field boundaries,

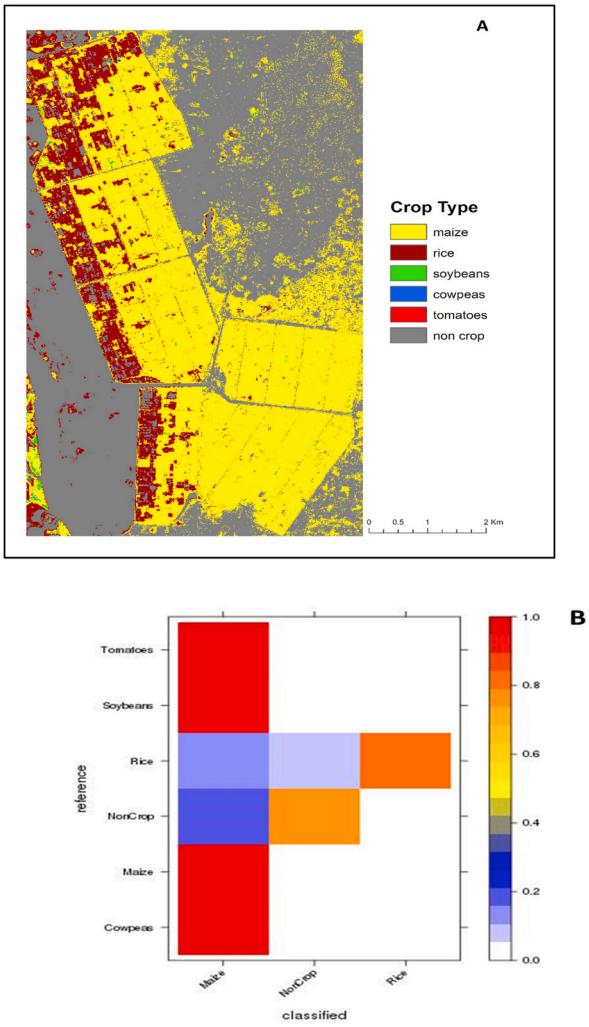


Fig. 3. a Rainy season classification for predominantly Rice (brown) and Maize (yellow). Figure 3b shows the confusion matrix for the rainy season calculated as the number of crops in each class divided by the total number of crops in the reference class and highlights the percentage of correct classifications (diagonal elements >70 percent) as well as the percentage of misclassifications (off diagonal elements). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

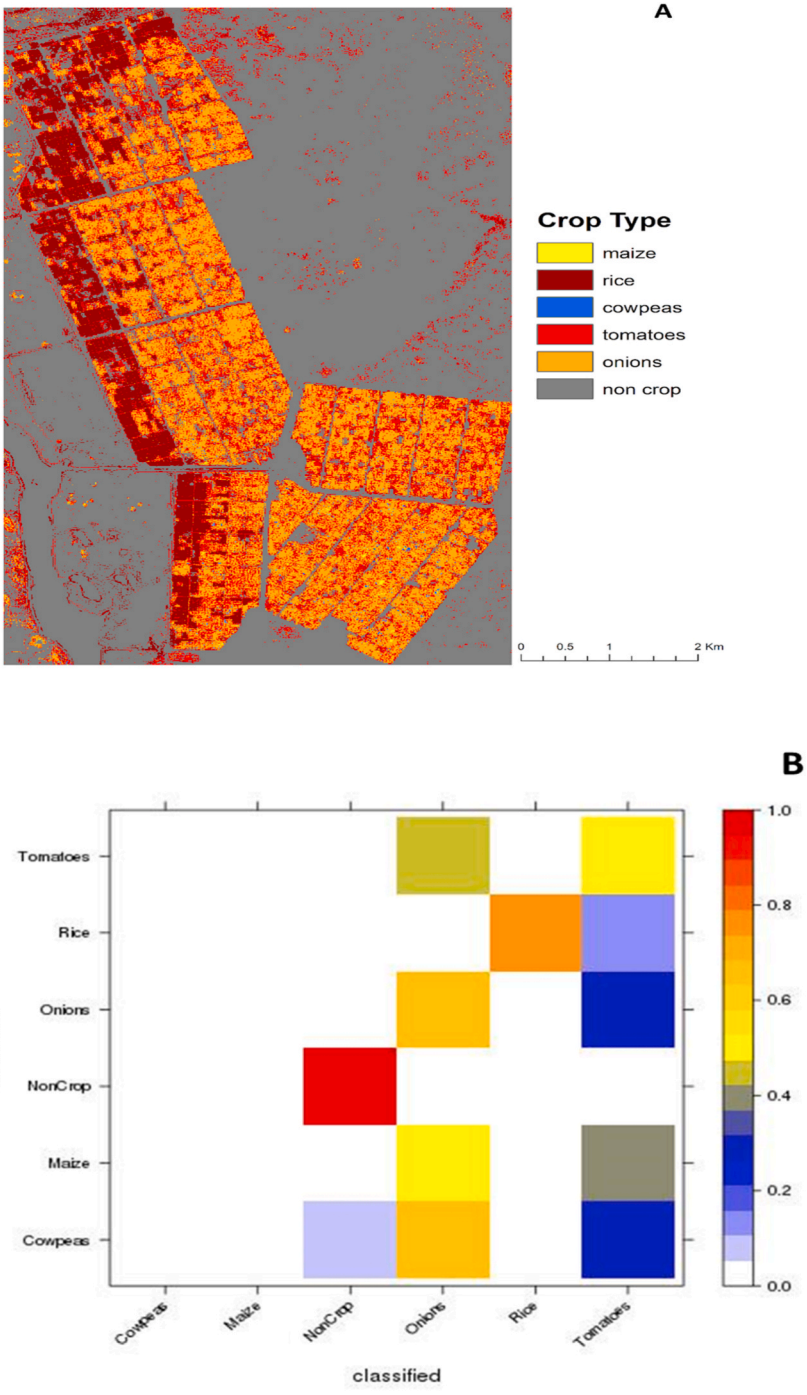


Fig. 4. a Dry season classification for predominantly Rice (brown) Onion (orange) and tomato (red). Figure 4b shows the confusion matrix classification and misclassification skill for the dry season calculated as the number of crops in each class divided by the total number of crops in the reference class and highlights the percentage crops placed into the correct class (diagonal elements >50 percent) as well as the percentage of time crops are incorrectly classified (off diagonal elements). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

increasing classification error. Of the three crop type classes, onions were the most distinguishable, placed into the correct class ~67% of the time and when misclassified were most often mistaken for tomatoes (~29%, Fig. 4b). When misclassified, Maize and tomatoes were most often mistaken for onion with a misclassification rate of approximately 53% and 44%, respectively. Maize was also misclassified as tomatoes at an equally high rate, 42%. Rice was classified with the greatest accuracy, and was placed into the correct class ~77% of the time (Fig. 4b). Similar to the rainy season, rates of misclassification were higher for undersampled crops in the dry season.

Cowpeas and maize comprise 14% of the in situ crop cut samples and were 100% omitted from their respective crop classes. When combined with small plot sizes and crops planted in close proximity within the Di perimeter, the degree of undersampling made crop identification challenging.

Random Forest crop type models with the greatest out-of-bag accuracies were used to assess crop composition. However, all factors from target season to target crop should be evaluated to select the best model. In light of that, the producer accuracy of each Random Forest crop type model was examined to ensure that the classification skill was maximized across all adequately sample crop types. For example, MEDMAXAMP model in the dry season exhibited the greatest overall skill for the well-sampled crops (rice, onions, and tomatoes). However, if classification skill of dry season rice was the primary interest, BAND7VI5 outperformed MEDMAXAMP with a 6% increase in rice classification; underscoring both model selection based on season and crops of interest but also the value of the data selected to train a Random Forest crop type model.

3.2. Estimating yield

Crop yield was estimated using Random Forest regression for crop types adequately sampled in both seasons. In the rainy season yield was estimated for rice and maize while rice, tomatoes, and onions were estimated in the dry season. A crop mask was generated from the most skillful crop classification results and applied to the final yield estimate to limit the yield evaluation to the targeted crop type. It is noted that yield was not reported for all crop cut samples and therefore the sample size is reduced from the totals reported in Table 1. In the rainy season there were 590 samples (maize: 462, rice: 128) and 615 samples in the dry (onions: 278, rice: 147, tomatoes: 190).

Similar to estimating crop composition, yield skill varied depending on the Random Forest model spectral inputs. In addition to variation in skill by season, there was variation in skill by crop type where some Random Forest models outperformed others when estimating yield, depending upon the crop of interest. Rainy season yield showed smaller changes in accuracy based on the yield model applied, where a ~5% difference in accuracy for rice and ~2% for maize was observed between the respective yield models (Fig. 2b). In the rainy season, BAND7VI5 was the most skillful yield model for both rice and maize though, maize yield accuracy was largely not impacted by the Random Forest regression model applied.

The greatest difference in yield accuracy occurred for dry season crops where variation in yield model skill was more dependent on the crop type. The BAND7VI5 model was most skillful for dry season rice yield similar to the most skillful model in the rainy season with an overall accuracy of 56%, and only a ~6% difference in accuracy observed between Random Forest models (Fig. 2c). MEDMAX and MEDMAXAMP models were most skillful for tomatoes and onions respectively with as much as an 81% difference between the most and least skillful tomato yield models.

Yield accuracy (computed from the MAPE) was greater with less error in the rainy season than the dry season. In the rainy season average estimated yields are consistent with average yield estimated from crop cuts (Ksoll et al., 2021) with accuracies greater than ~63% for rice and maize (Fig. 5). Yield estimates are within 0.2 t/ha of average crop cut estimates for rice and maize with MAE of less than 1.64 t/ha. In the dry season accuracies are greater than ~53% for rice and onions, with greater skill in estimating rice production than onions (MAE of 3.1 t/ha and 9.7 t/ha, for rice and onion).

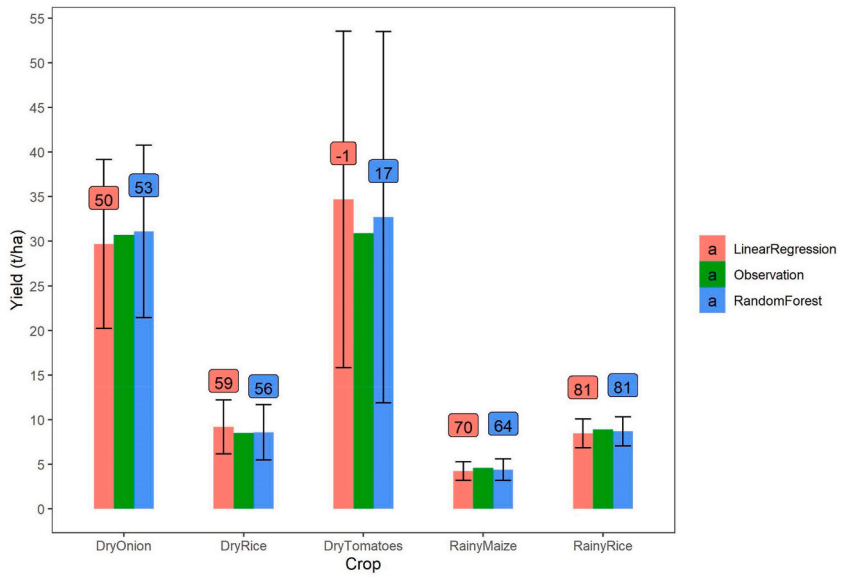


Fig. 5. Average yield by crop type measured from crop cut samples (green), estimated using Random Forest regression (blue) and linear regression (red) with corresponding measures of overall accuracy (computed from the MAPE, text boxes with corresponding colors). The error bars are a measure of the residual error as estimated from then Mean Absolute Error (MAE). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Dry season tomatoes exhibit low accuracy and large errors (MAE 20.1 t/ha) in yield estimation where the accuracy is ~17% indicating an inability to accurately estimate seasonal tomato production. In addition, the tomato yield models failed to converge with 100 decision trees where upwards of 400 trees were required to obtain yield model stability. When the tomato yield model finally stabilized with 400 decision trees the accuracy was ~10%. Though in situ tomato yields were collected throughout the growing season, using VI and green biomass-based approaches for estimating tomato yield may factor in the low accuracy achieved by the tomato yield model. The inclusion of additional biotic and abiotic stresses and the management of these stresses would be advantageous (Gosa et al., 2021).

3.3. Comparing linear and nonlinear methods of regression for yield estimation

The following section compares linear and nonlinear methods for estimating yield comparing ordinary least squares (OLS) regression and Random Forest regression. Rainy and dry season yield estimates are examined for each crop to identify advantages and limitations of each methodology. It is noted that slight differences in methodology occur between the linear and nonlinear methods, primarily the use of the S2 Top of Atmosphere (TOA) product and the inclusion of early season imagery in the time-series in the linear methodology. Due to the small size of the study domain, overall minimal differences (1%–3%) in skill were observed when using TOA instead of surface reflectance for nonlinear yield models (Figure A2.3). These differences in method do not significantly impact accuracy; therefore any differences in results can be attributed primarily to model type, linear or nonlinear, rather than model parameters and spectral inputs.

Evaluating both the MAE and MAPE linear regression multi-VI models out performed linear regression models built using a single vegetation index. Crops with lower average MAE (maize and rice) and higher accuracy (as computed from the MAPE) had minimal differences in linear regression model performance between the all-VI models and any single-VI analog. The med-max linear regression model for dry season rice, estimated plot-wise mean yield of 9.2 t/ha, and when incorporating all VIs and produced an MAE of 3.03 t/ha (estimated rainy season rice yields were 8.45 t/ha with MAE of 1.61 t/ha) (Table 4, Fig. 5). Comparatively, the Random Forest regression models estimated average dry season rice yields of 8.5 t/ha with an MAE of 3.12 t/ha (average rainy rice yields were 8.7 t/ha with MAE of 1.64 t/ha). Rainy season maize exhibits narrower differences between the all-VI linear regression model (1.02 t/ha) and the best single-VI linear regression models (1.07 t/ha for NDVI705), with an average forecasted yield of 4.24 t/ha. Similarly, the Random Forest regression model MAE was 1.22 t/ha and forecasted average maize yields of 4.7 t/ha (Fig. 5).

Across all crop types and seasons evaluated, rainy season rice and maize produced the highest accuracies in linear regression models (Table 5). The all-VI regression model for rice predicts yields with 81.2% accuracy in the rainy season and 59.5% accuracy in the dry season, comparable to the best performing Random Forest regression models (Fig. 5). Dry season tomato yields are predicted with the largest errors, with all linear regression models generating mean absolute prediction errors in excess of 100%. While tomatoes do have the highest average yield (34.7 t/ha), this value is not much greater than onions (29.7 t/ha), and yet tomato MAPE values are substantially larger than the MAPE values for dry season onion linear regression models (Fig. 5).

Overall, the nonlinear Random Forest and the linear med-max regression models produced comparable yield results across all rainy and dry season crop types indicating little difference between the model types. One advantage however is the transferability of Random Forest models, which is examined in section 3.4.

3.4. Interannual predictability of random forest models

In the following section, we apply all dry season 2019 Random Forest models to estimate crop type during dry season 2020 in effort to examine the predictive skill of a well-trained Random Forest crop type model. To assess skill, dry season 2020 survey data were used to derive accuracy and misclassification rates from the confusion matrix to evaluate prediction skill. In addition, new Random Forest crop type models are built with the dry season 2020 survey data to evaluate if a crop type model built with high quality crop cut data can predict future crop type as well as a crop type model built with survey data.

To evaluate the skill of a Random Forest crop type model calibrated with survey data, the 5 aforementioned models, which vary based on spectral input, were developed with dry season 2020 surveyed crop type. Similar to previous findings, the VI5 crop type model was the least skillful (out of bag accuracy 57.7%) while the MEDMAXAMP crop type model was the most skillful (out of bag accuracy 64.4%) (Fig. 6). Overall, comparable out of bag accuracy was achieved when building a Random Forest crop type model with either survey or crop cut data to estimate crop type for the time-period concurrent with the data collected.

The question however is the degree of interannual predictability from these data sources and if high quality crop cut data can make skillful predictions. To analyze this question, dry season 2019 Random Forest crop type models were used to estimate crop type for the 2020 dry season. Model estimated crop type was compared to a reference class derived from the 2020 dry season survey and rates of accuracy and misclassification were calculated to evaluate model interannual prediction skill. When the dry season 2019 Random

Table 4
Mean Absolute Error (t/ha) estimates from med-max linear regression crop yield models.

Season	Crop	N	All	GCVI	MTCI	NDVI705	NDVI740	NDVI
Dry	Onions	278	9.45	9.7	9.72	9.68	9.63	9.69
Dry	Rice	147	3.03	3.09	3.07	3.1	3.05	3.09
Dry	Tomatoes	190	18.85	19.65	20.07	19.82	20.14	19.92
Rainy	Maize	462	1.03	1.08	1.12	1.07	1.1	1.11
Rainy	Rice	128	1.61	1.79	1.8	1.82	1.8	1.81

Table 5
Accuracy (100-MAPE) (%) estimates from med-max linear regression crop yield models.

Season	Crop	N	All	GCVI	MTCI	NDVI705	NDVI740	NDVI
Dry	Onions	278	50.5	47.86	48.12	49.13	47.79	48.26
Dry	Rice	147	59.48	58.02	58.12	58.01	58.5	57.98
Dry	Tomatoes	190	−0.54	−11.65	−17.6	−16.06	−16.29	−16.07
Rainy	Maize	462	69.51	67.1	66.02	67.77	67	66.19
Rainy	Rice	128	81.22	79.05	78.91	78.71	78.88	78.90

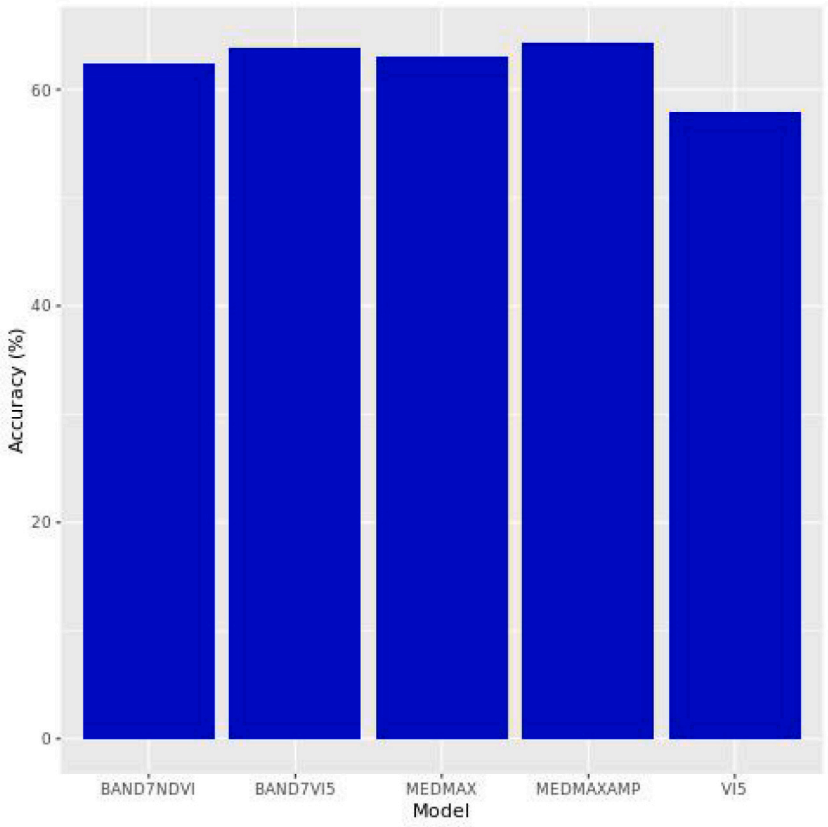


Fig. 6. The difference in model skill, out of bag accuracy, across 5 Random Forest models: BAND7VI5, BAND7NDVI, VI5, MEDMAXAMP, MEDMAX. Inputs for these models are from the dry season 2020 survey data collection for the Di perimeter in Burkina-Faso.

Forest crop type models were used to predict future crop type, predictive overall accuracy ranged from 52% (VI5 model) to 60% (MEDMAXAMP model) (Fig. 7).

Producer accuracy can be compared to determine how well each Random Forest crop type model performs. For this analysis, the 2019 MEDMAXAMP model is compared to the 2020 MEDMAXAMP model to evaluate the producer accuracy in estimating dry season 2020 crop type. It is noted that the dataset used to evaluate the 2020 MEDMAXAMP Random Forest model is the test subset of the 2020 dataset and therefore, the sample size is smaller. The producer accuracy is evaluated for the three well-sampled crops, including onions, rice and tomatoes. The 2020 MEDMAXAMP model accurately placed onions, rice and tomatoes into the correct class 51%, 25% and 31% of the time respectively (Fig. 8). Conversely, the producer accuracy of the 2019 MEDMAXAMP model is 49% (onions), 32% (Rice), and 50% (tomatoes). Overall, evaluating producer accuracy shows that Random Forest crop type models built with high quality crop cut data in 2019 are able to estimate future crop type in 2020 as skillfully as a model built with survey data from the concurrent year. It is noted that a high degree of producer accuracy was observed for the non-crop class and significantly contributes to the overall classification accuracy. Therefore, it is possible that a degree of skill is driven by the non-crop class considering features such as the water and bare ground are not changing on short interannual time scales.

Finally, to rule out spurious prediction skill, the dry season 2020 Random Forest crop type models are also applied backward in time to ensure the observed predictability is the result of high-quality crop cut data inputs into the 2019 Random Forest crop type models. When Random Forest crop type models trained with dry season 2020 survey data are used to estimate past crop type, in the dry season of 2019, the greatest overall accuracy was 50% for the MEDMAXAMP Random Forest model (Fig. 9). Highlighting the skill

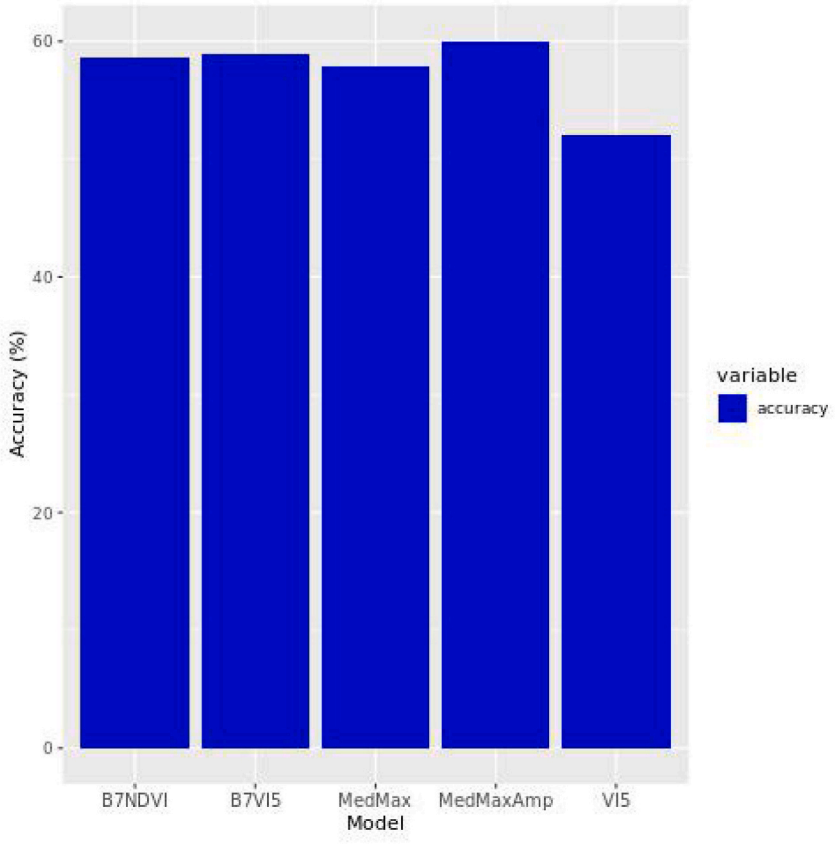


Fig. 7. The difference in model skill as rates of accuracy (blue), when Dry season 2019 models are used to estimate future crop type, in Dry season 2020. The metrics are derived from the number of crops correctly classified divided by the total. These metrics are computed for the 5 crop type models, BAND7VI5, BAND7NDVI, VI5, MEDMAXAMP and MEDMAX. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

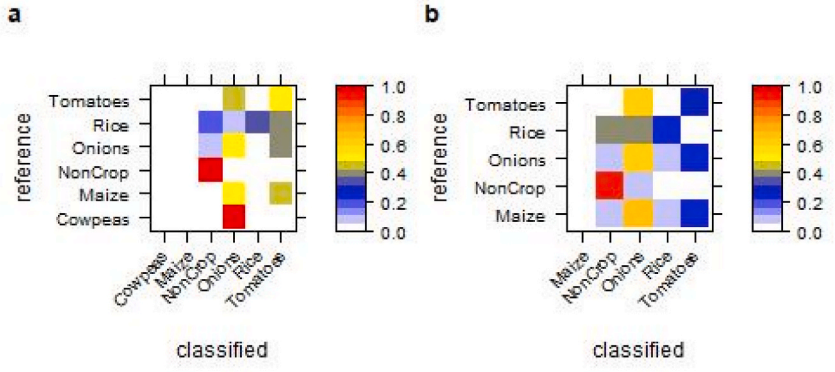


Fig. 8. The confusion matrix showing classification accuracy and misclassification estimating dry season 2020 crop types for the dry season MEDMAXAMP 2019 crop cut model (a) and the MEDMAXAMP 2020 survey data model (b). Skill is calculated as the number of crops in each class divided by the total number of crops in the reference class and highlights the percentage crops placed into the correct class (diagonal elements) as well as the percentage of time crops are incorrectly classified (off diagonal elements) emphasizing the comparable skill achieved predicting crop type at short interannual scales (skill in figure A compared to figure B).

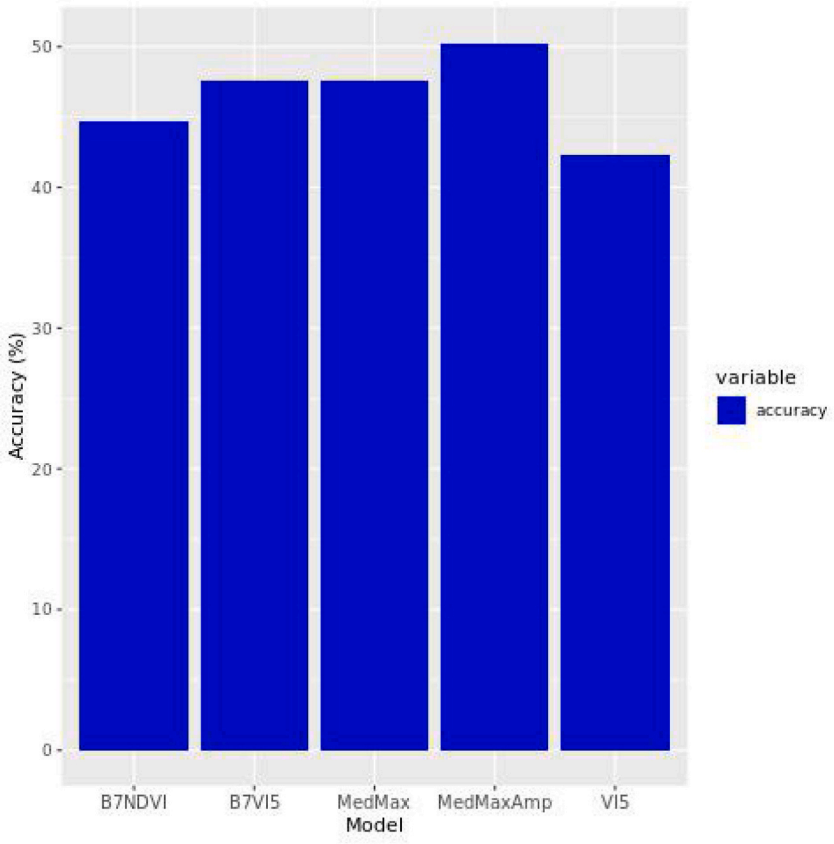


Fig. 9. Shows the difference in model skill as rates of accuracy (blue), when Dry season 2020 models are applied backwards in time and used to estimate dry season 2019 crop type. The metrics are derived from the number of crops correctly classified divided by the total. The metrics are derived for the 5 crop type models, BAND7VI5, BAND7NDVI, VI5, MEDMAXAMP and MEDMAX. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

attained, when predicting interannual crop type from survey data is less skillful than when estimating from higher quality crop cut data.

4. Discussion

4.1. Discussion

This study evaluated the skillfulness of Random Forest models calibrated with crop cut data for estimating crop type and yield in Burkina Faso and leveraged those Random Forest models to make interannual predictions of crop type using moderate resolution Sentinel-2. With skill comparable to other studies, in both the rainy and dry season (Burke and Lobell, 2017; Inglada et al., 2015; Lambert et al., 2018; Lobell et al., 2020), Sentinel-2 showed utility for remote evaluation in smallholder agricultural systems with landscape heterogeneity, particularly in the rainy season. Crop type and yield prediction skill decreases in the dry season as the proximity of crops grown within the same plot (Maize, Tomatoes, and Onions) pose a challenge for the coarser resolution of Sentinel-2. A similar finding to Lobell et al. (2020) who found intercropping decreased prediction skill unless there were larger differences between crop heights. In this case, fusion methods with LiDAR may prove useful distinguishing short from tall crops (Tommaso et al., 2021). High-Resolution imagery such as Planet Scope may further improve estimates (Rao et al., 2021). Consistent with previous findings (Belgiu and Drăguţ, 2016), under sampled crop types have higher commission and omission errors where as classes with better sampling and representation are favored by the Random Forest models. Methods, which include sparse data–limiting training data to well sampled classes in the Random Forest models – are potentially beneficial. Preliminary tests excluding under sampled classes had mixed results indicating that more research is needed. Future work will be conducted to test sparse data methods.

Evaluating feature importance identified several bands important for the Random Forest model's selection process. Overall, the bands in the red, red-edge and NIR ranges as well as vegetation indices built with these bands (MTCI, NDVI705, and NDVI740) emerged as important for the Random Forest models' selection process when estimating both crop type and yield across the best performing models (Appendix 3). However, variation existed among important features dependent upon target season and crop type. Models that included SWIR bands were optimal when identifying rainy season crop type or estimating rice yields in both seasons, where the SWIR bands were highly important features in those models. In addition, GCVI, which detects chlorophyll and canopy

greenness (Burke and Lobell 2017; Rao et al., 2021) was the most important feature in Maize yield models. Dry season skill showed improvement with crop type and yield model optimization where skill improved by varying the spectral inputs and metrics utilized by each Random Forest model. Therefore, it is important to select a model and its inputs based on the season and target crops of interest and consider platforms like Sentinel-2 which offer a broad band selection, advantageous for evaluating these landscapes. We also note that the crop cut collection objective was 6 focus crops and that other crops that may exist within the cultivated perimeter, specifically in the Dry season, introduce a degree of variability into the Random Forest models and may contribute to the decrease in Dry season skill.

Linear regression and nonlinear Random forest approaches for estimating yield were also evaluated in this study and were comparable in skill indicating these methods utilized were not a major determinant of prediction skill in a study area of this size. Instead, similar to Lobell et al. (2020), measurement or recall error on yield is a more likely determinant of yield model accuracy than the selected model types. It is noted that neither method evaluated in this study considered spatial or temporal dependence. Nonlinear methods, which consider spatial and temporal correlations, such as neural networks, might improve accuracy. In addition, harmonic regression may prove a more powerful linear regression model where accuracy increased as much as 7% and 12% across crop types in rainy and dry season yield estimates, respectively, when compared to the med-max linear and nonlinear Random Forest regression models (Appendix 4). Similarly, decreases in MAE were also observed across all crop types, though the accuracy and error of rice yield was largely unchanged across methodologies.

This study also evaluated the crop type prediction skill of Random Forest crop type models calibrated with survey data and found comparable out of bag accuracies between crop cut and survey data calibrated models. However, producer accuracy was significantly higher for crop type models calibrated with crop cut data. Generally larger bias and as a result reduced prediction accuracy is expected in survey data calibrated models though some studies suggest reducing GPS error can improve data quality (Burke and Lobell, 2017). Therefore, the reduction in producer accuracy could be indicative of survey bias and geolocation error. However, with consideration of the geolocation error (5–9m) in the 2020 survey data used in this study, the reduction in accuracy could also be indicative of a sampling issue (Table 1 & Appendix 5). The sample size within the perimeter in the 2020 survey was 56%–72% percent lower for onions and rice than the 2019 crop cut sample respectively. Future work should evaluate an expanded crop type sample size, within the perimeter and neighboring fields, to examine the impact on producer accuracy as field data collection is costly and accurately geo-located survey data could offer potentially cost-effective methods of data collection.

Finally, this study evaluated the skillfulness of crop type models calibrated with high quality crop cut data for predicting interannual crop type. Overall Random Forest crop type models calibrated with 2019 crop cut data and leveraged to predict interannual crop type performed as well as Random Forest crop type models calibrated with 2020 survey data to predict Dry season 2020 crop type. Specifically, the 2019 crop cut calibrated Random Forest crop type models used for interannual predictions exhibited similar producer accuracy to the 2020 survey data calibrated models. Crops like rice were more accurately classified in the 2019 crop type models than the 2020 crop type models. Therefore, in absence of data, leveraging a well-trained crop type model is valuable for estimating interannual crop type. However, it is unclear how far into the future this prediction skill extends. In addition, the data used for calibration is essential for the prediction performance of a Random Forest crop type model leveraged for uncalibrated interannual predictions, where those trained with survey data and leveraged for interannual prediction performed poorly when compared to those trained with crop cut data.

4.2. Limitations

The study area was comprised of multiple crop types in addition to the 6 focus crops sampled during the crop cut surveys. Therefore, not all crops within the Di perimeter were sampled, increasing inter and intra plot variability that when combined with S2 spatial resolution, limit crop type and yield estimation accuracy. Additionally, tomato and onion crops are radiation and moisture dependent; so spectral information and green biomass specific vegetation indices alone may be inadequate to fully capture yield for these crops. Future work should employ climate information for the tomato and onion yield models, particularly as these are high value crops. Finally, interannual predictive crop type models are calibrated using crop type data from one season and applied to the following season. Future work should expand the in-situ data used to calibrate the models as well as examine how far into the future the crop type models can make skillful predictions to understand the capabilities of this uncalibrated approach.

5. Conclusions

Earth observing systems are a powerful tool for monitoring and evaluating agriculture and when combined with machine learning can provide skillful measures of crop type and yield mapping across seasons and landscapes. Moderate spatial resolution Sentinel-2 combined with a crop cut calibrated Random Forest model are skillful predictors of crop type and rice and maize yield in smallholder agricultural systems with landscape heterogeneity. We find the resolution of Sentinel-2 is adequate resolution for remotely detecting crop conditions particularly because of the large number of bands offered, where the red through NIR bands, SWIR bands and GCVI were pivotal spectral inputs. The calibrated Random Forest models can be leveraged to make skillful predictions of agricultural productivity and ultimately food availability of nearby communities and for years with no training data. Given increasing global food prices and restricted commodity trade, understanding local agricultural productivity using affordable and timely remote sensing-based methods is essential for ensuring appropriate humanitarian interventions.

Author contributions

All authors meet authorship criteria and have participated in manuscript concept, design, methodology, writing, analysis, or

revisions. Conceptualization and design: A.E., M.L.C., and C.S.R.N.; Methodology and formal analysis: A.E., M.L.C., C.S.R.N., A.L.D., C. K., and M.R.W.; original manuscript preparation: A.E.; Writing, Editing, Review and Revisions: A.E., M.L.C., C.S.R.N., A.L.D., C.K., M. R.W., and M.E.B.; Visualization: A.E., M.R.W., and A.L.D; Data collection: A.L.D. and C.K.; Food security expertise and consult: M.E.B.; Project management: C.S.R.N and M.L.C.; Funding acquisition: C.S.R.N. and M.L.C.; Project administration: A.E.; Supervision: M.L.C. and C.S.R.N. All authors have read and agreed to the published version of the manuscript.

Ethical statement

Along with my Coauthors, I declare that all ethical practices in relation to the development, writing and publishing of this manuscript have been reviewed and followed.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Akiko Elders reports financial support was provided by Millennium Challenge Corporation. Mark L Carroll reports financial support was provided by Millennium Challenge Corporation. Christopher S.R. Neigh reports was provided by Millennium Challenge Corporation. Anthony Louis D'Agostino reports financial support was provided by Millennium Challenge Corporation. Christopher Ksoll reports financial support was provided by Millennium Challenge Corporation. Margaret Wooten reports financial support was provided by Millennium Challenge Corporation. The authors are funded for this work by Millennium Challenge Corporation (an autonomous US Federal Government Agency) who empower smallholder farmers in developing nations through Agricultural Development Projects.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was supported by Millennium Challenge Corporation. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views of this agency. We thank John Molyneaux and Hamissou Samari from Millennium Challenge Corporation for their contributions to this work. We thank the peer reviewers for their comments and suggestions.

Appendix 1

I. 2018–2019 Crop Cut Data Collection

The crop-cut survey proceeded as follows. After contacting the selected households, interviewers first collected some information on the plot owner and the type of crops that had been planted. If the plot was planted with more than one crop, was very large, or had differing productivity within the field, interviewers subdivided the plot into subplots for separate crop cuts. Enumerators then paced the outline of each subplot with the GPS-enabled tablet, recording the outline of the plot. After the interviewer captured the subplot's outline, a computer program directed the interviewer to a random location for the measurement square in the subplot. There the interviewer placed 5 m × 5 m measurement squares—four poles with a string attached between them—in the ground to mark the area in which measurements were to be taken. Between the placement of the measurement squares and harvesting, interviewers regularly returned to the fields to observe whether measurement squares were still in the initial place and strings were tightly strung. At harvest time, interviewers returned to harvest and weigh the crops within the measurement square. The geo-coded data on crop type and yields is used to provide ground-truthing information for the remote sensing effort, as well as information on yields.

Our in-situ data were collected at the level of 5 m × 5 m crop cut measurement squares, and include the crop type, planting and harvesting dates, and yield data. The accuracy of the location information collected through GPS devices is bounded by the device's location measurement error, which for our readings usually range between 2 and 6 m, though larger errors did occur. As a result, measurement squares do not retain the true appearance of a square. Self-reported parcel-specific planting and harvesting dates informed our decision of the season windows to use. The enumeration team recorded GPS measurements using a program called "parcel mapper" on a tablet computer (Samsung Galaxy Tab A6 T285) for a subset of plots on the Di perimeter, as well as measurement squares that were the basis of conducting rainy-season and dry season crop cuts associated with those plots. Measurement squares delimited the area from which crops were to be harvested and then weighed, to provide a more accurate estimate of crop productivity than self-reported data provide (Lobell et al., 2020). For each selected square, yield data were recorded once for crops with a single harvest and multiple times for crops with possible multiple rounds of harvesting, as in the case of crops like tomatoes. Crops like tomatoes are routinely harvested multiple times in a season, and therefore total output is the sum of harvest weights across all harvest events. The clean version of these data includes the plot boundaries (739 dry season plots and 660 rainy season plots), measurement square centroids (911 dry season squares and 613 rainy season squares), details on the crop type and total yield, and a unique identifier to ensure that all subsequent outputs can be merged back to the original data. Crop yield data were not available for all parcels, and therefore analyses requiring yield data were based on fewer observations than analyses that only required crop type information. Table 1 describes the in situ data used for the two seasons of the analysis.

II. 2020 Survey Data Collection

The survey data collection targeted landowners and renters with land in the perimeter (who include winners of a land lottery, the Di lottery) as well as applicants to the Di Lottery who did not win the lottery. This survey had initially intended to collect information on crop type from a sample of over 2000 households. Because of terrorist activity in the area, the in-person survey, which collected geo-located crop type, was switched to an abbreviated survey by telephone without geo-localized crop type data. As a result, the scope of data collected was dramatically reduced and there is very little overlap in the two in situ crop-type samples. While this data format did result in collecting (some) crop type information, high-quality crop yield data was not collected.

Appendix 2



Figure A2.1. The change in out of bag accuracy (green) and processing time (blue) as the number of decision trees increases from 10 to 300.

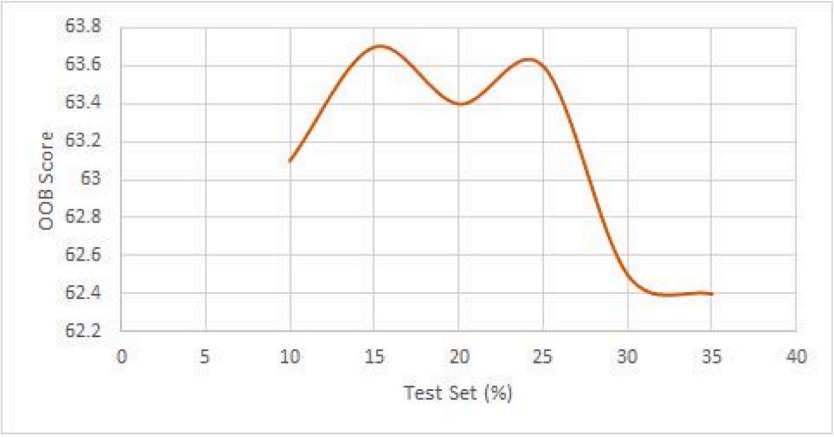


Figure A2.2. The change in accuracy with change in the percentage of data reserved for validation.

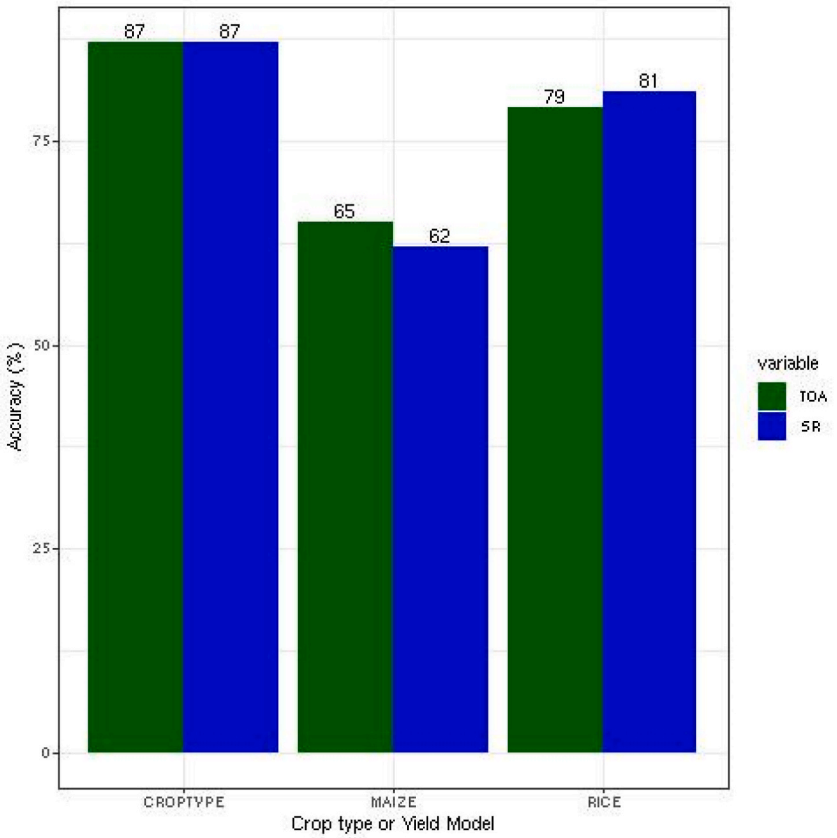


Figure A2.3. The change in accuracy of rainy season BAND7NDVI model comparing the Top of Atmosphere (Green) and Surface Reflectance (blue) Sentinel-2 products.

Appendix 3

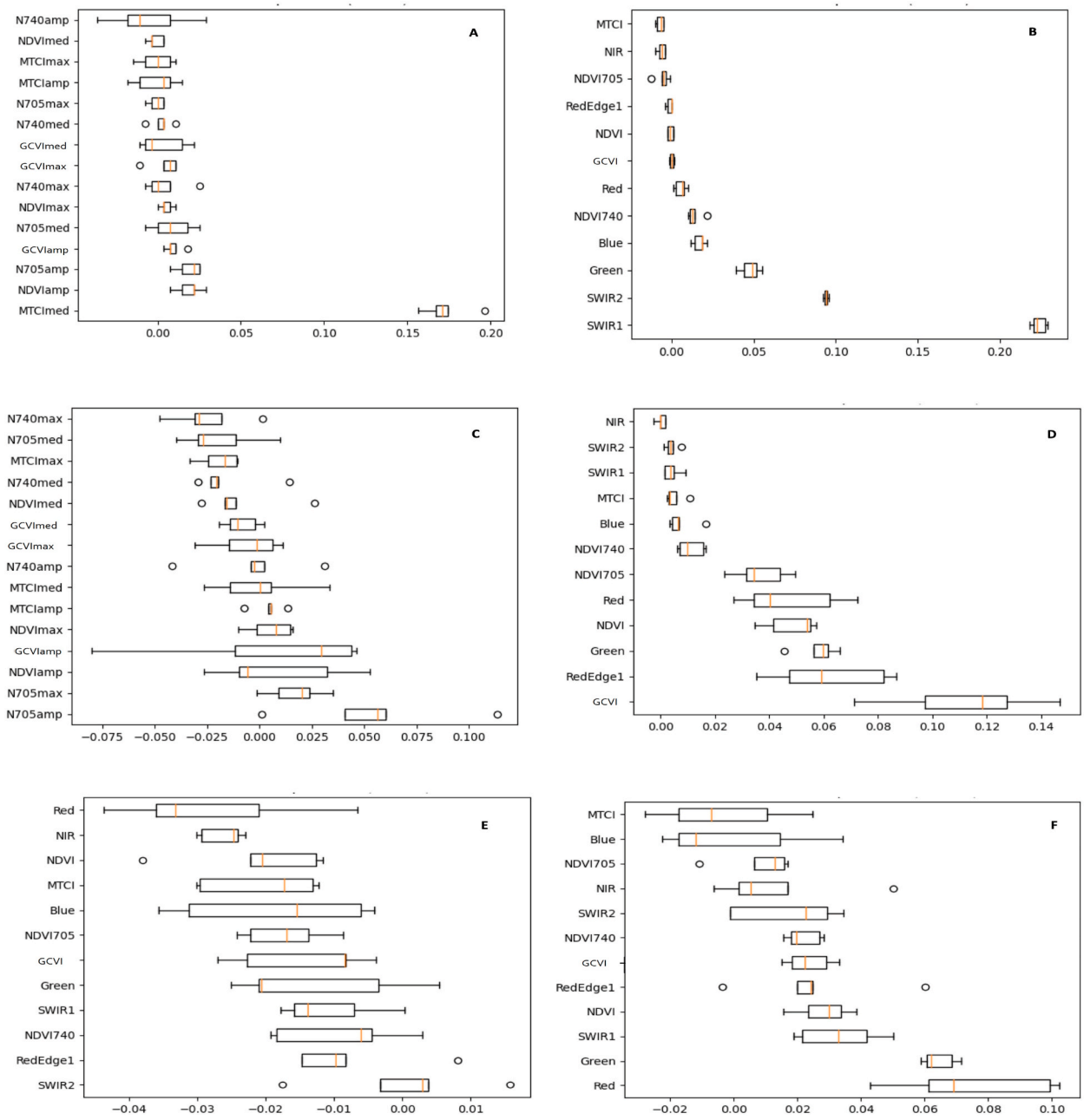


Figure A.3.1. The importance of each spectral band as computed by Permutation feature importance for the Dry (a) and Rainy (b) season crop type and yield for dry season onions (c), rainy season maize (d), dry season rice (e), and rainy season rice (f).

Appendix 4

Harmonic regressions reduce data dimensionality by translating seasonal dynamics into a small number of features (Wang et al., 2020; Wilson et al., 2018), as shown in Equation A4.1 which estimates a linear time trend, four sinusoidal parameters, and a constant for each crop observation as a function of a single spectral band or vegetation index.

$$f(t) = \alpha t + \sum_{j=1}^{2} \left(a_j cos(2\pi j \omega t) + b_j sin(2\pi j \omega t) \right) + C \quad \text{(A4.1)}$$

The recovered parameter estimates from Equation A4.1 are covariates for estimating yields (Equation A4.2), and our specification incorporates the harmonic regression estimates for V number of spectral bands and/or vegetation indices. Figure A4.1 illustrates how the harmonic regression approach with just a small number of estimated parameters can approximate the observed time-series of an

individual vegetation index over the course of a season.

$$y = \beta_0 + \sum_{v=1}^{V}\left(\beta_1{}^{v}\widehat{\alpha^{v}} + \sum_{j=1}^{2}\left(\beta_{2,j}{}^{v}\widehat{a_j^{v}} + \beta_{3,j}{}^{v}\widehat{b_j^{v}}\right)\right) + \varepsilon \tag{A4.2}$$

In all regression models, *y,* the log yield (tons/hectare), is calculated by dividing the weighted output measured from the crop cut activities by the measurement square area (5m × 5m). Predicted yield, MAE, and MAPE values are derived by exponentiating the results back to yields in levels.

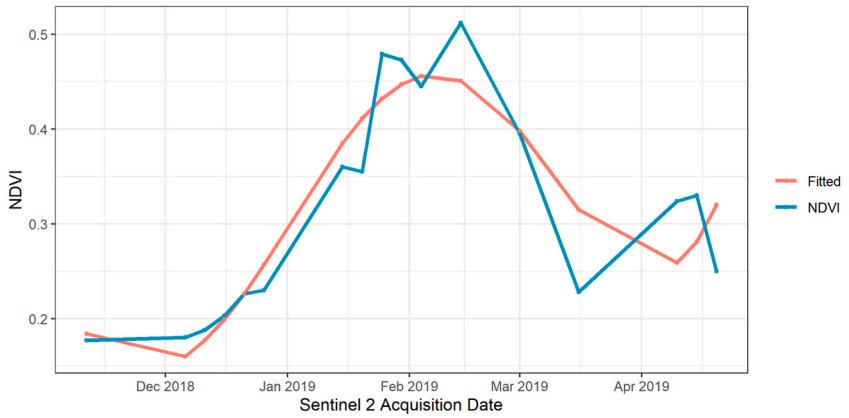


Figure A4.1. Harmonic regression to model seasonal VI dynamics. Results correspond to a single location (−3.413, 13.2297). Fitted results are derived from a two-cycle harmonic regression model for the dry season spanning all available Sentinel-2 images for November 1, 2018–April 30, 2019.

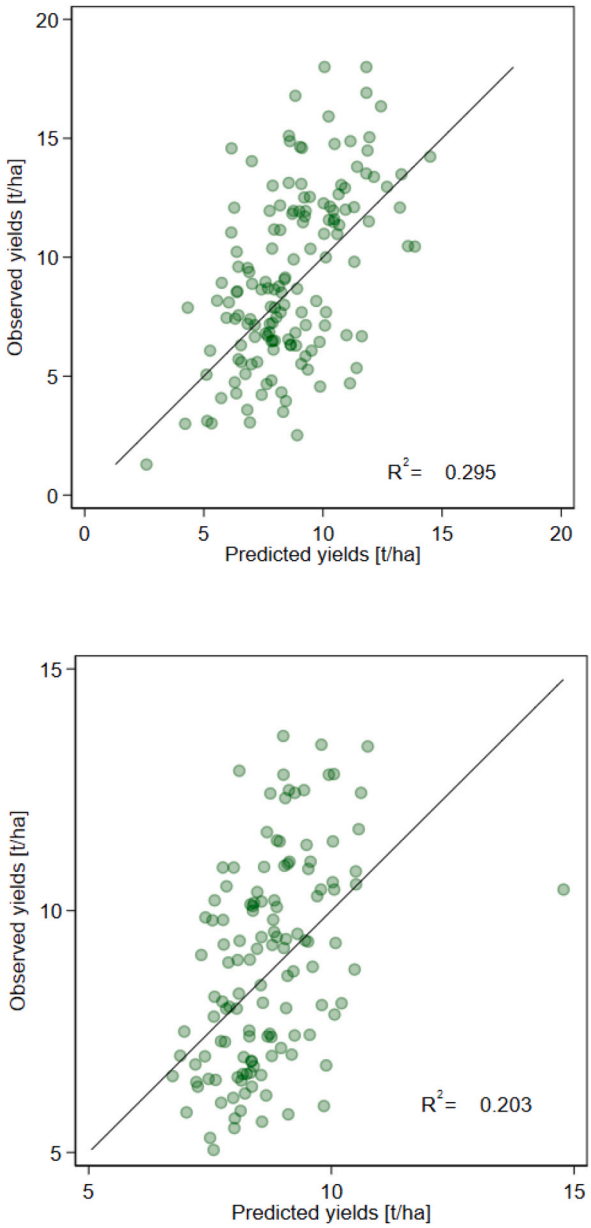


Figure A4.2. Comparison of observed and predicted rice yields using saturated VI harmonics regression model for the Dry (top) and rainy (bottom) season. Results are based on regressing log rice yields on the full set of vegetation index harmonic regression coefficients on the full sample of data and then exponentiating predicted values to return predicted yields in levels.

Table A4.1
MAE (t/ha) estimates from linear regression crop yield models

Season	Crop	N	Vegetation Index Harmonic Regression Models					
			All	GCVI	MTCH	NDVI705	NDVI740	NDVI
Dry	Onions	278	8.74	9.48	9.72	9.37	9.33	9.39
Dry	Rice	147	2.49	3.03	3.06	3.08	3.03	3.08
Dry	Tomatoes	190	17.32	18.76	20.05	18.85	19.13	18.89
Rainy	Maize	462	0.98	1.03	1.12	1.02	1.04	1.04
Rainy	Rice	128	1.60	1.74	1.82	1.76	1.76	1.76

Table A4.2
Accuracy (100-MAPE) (%) estimates from linear regression crop yield models

Season	Crop	N	Vegetation Index Harmonic Regression Models					
			All	GCVI	MTCI	NDVI705	NDVI740	NDVI
Dry	Onions	278	58.13	51.78	48.07	51.56	52.88	52.19
Dry	Rice	147	67.55	59.37	58.35	58.62	59.9	58.89
Dry	Tomatoes	190	9.99	−7.36	−16.71	−8.24	−6.64	−8.9
Rainy	Maize	462	71.41	68.87	65.87	69.21	68.75	68.17
Rainy	Rice	128	81.46	79.7	78.6	79.4	79.52	79.41

Appendix 5

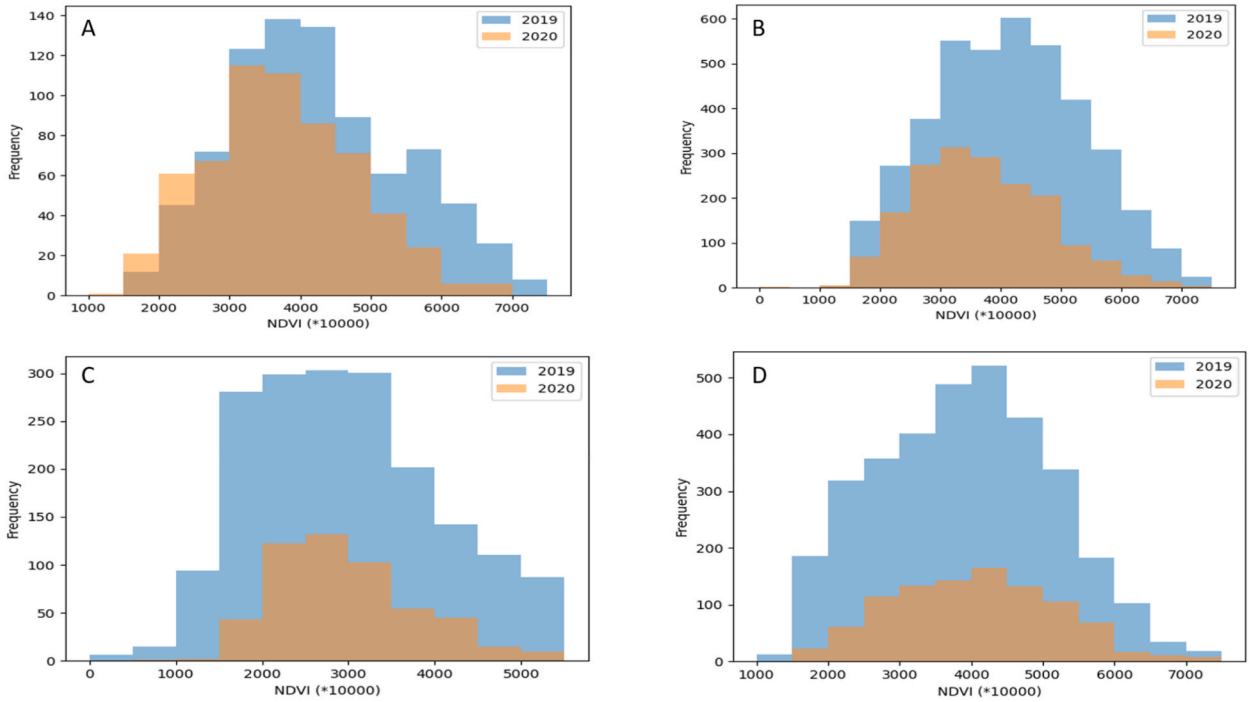


Figure A.5.1. NDVI distribution for crop types (maize (a), onions (b), rice (c), and tomatoes (d)) in the 2019 crop cut and 2020 survey collections. Though the number of samples in the 2019 collection is greater, the distribution of NDVI is similar between years showing consistency in the crop's spectral signature between collections, consistent with the higher degree of geolocation accuracy in the 2020 survey data collection effort.

References

Abubakar, G.A., Wang, K., Shahtahamssebi, A., Xue, X., Belete, M., Gudo, A.J.A., Mohamed Shuka, K.A., Gan, M., 2020. Mapping maize fields by using multi-temporal sentinel-1A and sentinel-2A images in makarfi, northern Nigeria, Africa. Sustainability 12, 2539. https://doi.org/10.3390/su12062539.

Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., A.de By, R., 2018. A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. Rem. Sens. 10, 729. https://doi.org/10.3390/rs10050729.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Rem. Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, M.E., Kshirsagar, V., 2015. Weather and International Price Shocks on Food Prices in the Developing World, vol. 35. Glob Environ Chang.

Brown, M.E., Funk, C.C., Galu, G., Choularton, R., 2007. Earlier famine warning possible using remote sensing and models. Eos Trans. Am. Geophys. Union 88, 381–382. https://doi.org/10.1029/2007EO390001.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. USA 114, 2189–2194. https://doi.org/10.1073/pnas.1616919114.

Carletto, C., Gourlay, S., Winters, P., 2015. From guesstimates to GPStimates: land area measurement and implications for agricultural analysis. J. Afr. Econ. 24, 593–628. https://doi.org/10.1093/jae/ejv011.

Carletto, C., Corral, P., Guelfi, A., 2017. Agricultural commercialization and nutrition revisited: empirical evidence from three African countries. Food Pol. 67, 106–118. https://doi.org/10.1016/j.foodpol.2016.09.020.

Carroll, M.L., DiMiceli, C.M., Townshend, J.R.G., Sohlberg, R.A., Elders, A.I., Devadiga, S., Sayer, A.M., Levy, R.C., 2017. Development of an operational land water mask for MODIS Collection 6, and influence on downstream data products. Int. J. Digit. Earth 10, 207–218. https://doi.org/10.1080/17538947.2016.1232756.

Chang, J., Hansen, M.C., Pittman, K., Carroll, M., DiMiceli, C., 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron. J. 99, 1654–1664. https://doi.org/10.2134/agronj2007.0170.

Chellasamy, M., Ferré, T.P.A., Greeve, M.H., 2014. Automatic training sample selection for a multi-evidence based crop classification approach. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch. 40, 63–69. https://doi.org/10.5194/isprsarchives-XL-7-63-2014.

Conrad, C., Dech, S., Dubovyk, O., Fritsch, S., Klein, D., Löw, F., Schorcht, G., Zeidler, J., 2014. Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. Comput. Electron. Agric. 103, 63–74. https://doi.org/10.1016/j.compag.2014.02.003.

Debats, S.R., Luo, D., Estes, L.D., Fuchs, T.J., Caylor, K.K., 2016. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens. Environ. Times 179, 210–221. https://doi.org/10.1016/j.rse.2016.03.010.

dela Torre, D.M.G., Gao, J., Macinnis-Ng, C., 2021. Remote sensing-based estimation of rice yields using various models: a critical review. Geo Spatial Inf. Sci. 24, 580–603. https://doi.org/10.1080/10095020.2021.1936656.

Doraiswamy, P., 2004. Crop condition and yield simulations using Landsat and MODIS. Remote Sens. Environ. Times 92, 548–559. https://doi.org/10.1016/j.rse.2004.05.017.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sens. Environ. Times 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026.

Esri, 2009. "World" [basemap]. Scale Not Given. "World Imagery". December 12. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9.

European Space Agency, 2016. Sen2Cor 2.2.1—Software Release Note. Paris, France: European Space Agency. Paris. France.

Fan, J., Defourny, P., Zhang, X., Dong, Q., Wang, L., Qin, Z., De Vroey, M., Zhao, C., 2021a. Crop mapping with combined use of European and Chinese satellite data. Rem. Sens. 13, 4641. https://doi.org/10.3390/rs13224641.

Fan, J., Zhang, X., Zhao, C., Qin, Z., De Vroey, M., Defourny, P., 2021b. Evaluation of crop type classification with different high resolution satellite data sources. Rem. Sens. 13, 911. https://doi.org/10.3390/rs13050911.

FAO, 2021. The State of Food and Agriculture 2021. Making Agrifood Systems More Resilient to Shocks and Stresses. FAO, Rome. https://doi.org/10.4060/cb4476en.

Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu, J., Hansen, M., Obersteiner, M., 2011. Cropland for sub-Saharan Africa: a synergistic approach using five land cover data sets. Geophys. Res. Lett. 38 https://doi.org/10.1029/2010GL046213.

Gerstmann, H., Möller, M., Gläßer, C., 2016. Optimization of spectral indices and long-term separability analysis for classification of cereal crops using multi-spectral RapidEye imagery. Int. J. Appl. Earth Obs. Geoinformation 52, 115–125. https://doi.org/10.1016/j.jag.2016.06.001.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. Times 202, 18–27.

Gosa, S.C., Koch, A., Shenhar, I., Hirschberg, J., Zamir, D., Moshelion, M., 2021. Predicting tomato field-yield using continuous monitoring of young tomato water status. https://doi.org/10.1101/2021.03.18.435447.

Grace, K., Nagle, N.N., Husak, G., 2016. Can small-scale Agricultural production improve children's health? Examining stunting vulnerability among very young children in Mali, West Africa. Ann. Assoc. Am. Geogr. 106, 722–737. https://doi.org/10.1080/24694452.2015.1123602.

Grace, K., Frederick, L., Brown, M.E., Boukerrou, L., Lloyd, B., 2017. Investigating important interactions between water and food security for child health in Burkina Faso. Popul. Environ. 39, 26–46. https://doi.org/10.1007/s11111-017-0270-6.

Human Right Watch World Report, 2022. Burkina Faso. New York NY: HRW. Available at: https://www.hrw.org/world-report/2022/country-chapters/burkina-faso.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Rem. Sens. 7, 12356–12379. https://doi.org/10.3390/rs70912356.

Kim, H.-O., Yeom, J.-M., 2014. Effect of red-edge and texture features for object-based paddy rice crop classification using RapidEye multi-spectral satellite image data. Int. J. Rem. Sens. 1–23. https://doi.org/10.1080/01431161.2014.965285.

Ksoll, Christopher, Morgan, Seth, Kristine, Bos, Randall, Blair, April 2018. Evaluation of the Burkina Faso Agriculture Development Project: Baseline Report. Mathematica, Washington, DC.

Ksoll, Christopher, Randall, Blair, Morgan, Seth, Yiriyibin, Bambio, Lauver, Caroline, July 2019. Evaluation of the Burkina Faso Agriculture Development Project: Interim Report. Mathematica, Washington, DC.

Ksoll, Christopher, Moroz, Elena, Turiansky, Abbie, D'Agostino, Anthony, Bambio, Yiriyibin, Roberge, Daniel, Blair, Randall, April 2021. Evaluation of the Burkina Faso Agriculture Development Project: Final Report. Mathematica, Washington, DC.

Labus, M.P., Nielsen, G.A., Lawrence, R.L., Engel, R., Long, D.S., 2002. Wheat yield estimates using multi-temporal NDVI satellite imagery. Int. J. Rem. Sens. 23, 4169–4180. https://doi.org/10.1080/01431160110107653.

Lambert, M.-J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sens. Environ. Times 216, 647–657. https://doi.org/10.1016/j.rse.2018.06.036.

Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I. ul H., Singha, K., Choudhury, S.B., Chetri, T., Malek, Ž., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturn, T., Karner, M., McCallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., Fritz, S., 2019. Estimating the global distribution of field size using crowdsourcing. Global Change Biol. 25, 174–186. https://doi.org/10.1111/gcb.14492.

Lobell, D.B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., Murray, S., 2020. Eyes in the sky, boots on the ground: assessing satellite- and ground-based approaches to crop yield measurement and analysis. Am. J. Agric. Econ. 102, 202–219. https://doi.org/10.1093/ajae/aaz051.

Lowder, Sarah K., Jakob, Skoet, Terri, Raney, 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World Dev. 87, 16–29.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogrammetry Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Rem. Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343.

McCarty, J.L., Neigh, C.S.R., Carroll, M.L., Wooten, M.R., 2017. Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery. Remote Sens. Environ. Times 202, 142–151. https://doi.org/10.1016/j.rse.2017.06.040.

McNally, A., Husak, G.J., Brown, M., Carroll, M., Funk, C., Yatheendradas, S., Arsenault, K., Peters-Lidard, C., Verdin, J.P., 2015. Calculating crop water requirement satisfaction in the West Africa Sahel with remotely sensed soil moisture. J. Hydrometeorol. 16, 295–305. https://doi.org/10.1175/JHM-D-14-0049.1.

Müller-Wilm, U., 2016. Sentinel-2 MSI—Level-2a Prototype Processor Installation and User Manual. Available online. http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf.

Neigh, Christopher S.R., Carroll, Mark L., Wooten, Margaret R., McCarty, Jessica L., Powell, Bristol F., Husak, Gregory J., Enenkel, Markus, Hain Christopher, R., 2018. Smallholder crop area mapped with wall-to-wall WorldView sub-meter panchromatic image texture: a test case for Tigray, Ethiopia. Rem. Sens. Environ. 212, 8–20. https://doi.org/10.1016/j.rse.2018.04.025.

Niles, M.T., Brown, M.E., 2017. A multi-country assessment of factors related to smallholder food security in varying rainfall conditions. Sci. Rep. 7, 16277. https://doi.org/10.1038/s41598-017-16282-9.

Palanivel, K., Surianarayanan, C., 2019. AN APPROACH FOR PREDICTION OF CROP YIELD USING MACHINE LEARNING AND BIG DATA TECHNIQUES. IJCET 10. https://doi.org/10.34218/IJCET.10.3.2019.013.

Palchowdhuri, Y., Valcarce-Diñeiro, R., King, P., Sanabria-Soto, M., 2018. Classification of multi-temporal spectral indices for crop type mapping: a case study in Coalville, UK. J. Agric. Sci. 156, 24–36. https://doi.org/10.1017/S0021859617000879.

Quarmby, N.A., Milnes, M., Hindle, T.L., Silleos, N., 1993. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. Int. J. Rem. Sens. 14, 199–210. https://doi.org/10.1080/01431169308904332.

Rao, P., Zhou, W., Bhattarai, N., Srivastava, A.K., Singh, B., Poonia, S., Lobell, D.B., Jain, M., 2021. Using sentinel-1, sentinel-2, and Planet imagery to map crop type of smallholder farms. Rem. Sens. 13, 1870. https://doi.org/10.3390/rs13101870.

Rockström, J., Barron, J., Fox, P., 2003. Water productivity in rain-fed agriculture: challenges and opportunities for smallholder farmers in drought-prone tropical agroecosystems. In: Kijne, J.W., Barker, R., Molden, D. (Eds.), Water Productivity in Agriculture: Limits and Opportunities for Improvement. CABI, Wallingford, pp. 145–162. https://doi.org/10.1079/9780851996691.0145.

Schut, A.G.T., Traore, P.C.S., Blaes, X., de By, R.A., 2018. Assessing yield and fertilizer response in heterogeneous smallholder fields with UAVs and satellites. Field Crop. Res. 221, 98–107. https://doi.org/10.1016/j.fcr.2018.02.018.

Scikit-learn, 2011. Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825–2830.

Son, N.T., Chen, C.F., Chen, C.R., Minh, V.Q., Trung, N.H., 2014. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. Agric. For. Meteorol. 197, 52–64. https://doi.org/10.1016/j.agrformet.2014.06.007.

Tommaso, S.D., Wang, S., Lobell, D.B., 2021. Combining GEDI and Sentinel-2 for wall-to-wall mapping of tall and short crops. Environ. Res. Lett. 16, 125002 https://doi.org/10.1088/1748-9326/ac358c.

Turker, M., Ozdarici, A., 2011. Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study. Int. J. Rem. Sens. 32, 9735–9768. https://doi.org/10.1080/01431161.2011.576710.

Turner, M.D., Teague, M., Ayantunde, A., 2021. Livelihood, culture and patterns of food consumption in rural Burkina Faso. Food Secur. 13, 1193–1213. https://doi.org/10.1007/s12571-021-01150-2.

Ursani, A.A., Kpalma, K., Lelong, C.C.D., Ronsin, J., 2012. Fusion of textural and spectral information for tree crop and other agricultural cover mapping with very-high resolution satellite images. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 5, 225–235. https://doi.org/10.1109/JSTARS.2011.2170289.

USAID Burkina Faso, 2015. Fact Sheet: Agriculture and Food Security. Washington DC: USAID, vol. 2015. Available at: https://www.usaid.gov/sites/default/files/documents/1860/BF%20Fact%20Sheet%20-%20Ag%20%26%20FS%201215.pdf.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., 2015. Processing Sentinel-2 image time series for developing a real-time cropland mask, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). In: Presented at the IGARSS 2015-2015 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Milan, Italy, pp. 2731–2734. https://doi.org/10.1109/IGARSS.2015.7326378.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Rem. Sens. 8, 55. https://doi.org/10.3390/rs8010055.

van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: a systematic literature review. Comput. Electron. Agric. 177, 105709 https://doi.org/10.1016/j.compag.2020.105709.

Verdin, J., Funk, C., Senay, G., Choularton, R., 2005. Climate science and famine early warning. Phil. Trans. Biol. Sci. 360, 2155–2168. https://doi.org/10.1098/rstb.2005.1754.

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.-T., 2018. How much does multi-temporal Sentinel-2 data improve crop type classification? Int. J. Appl. Earth Obs. Geoinformation 72, 122–130. https://doi.org/10.1016/j.jag.2018.06.007.

Waldhoff, G., Lussem, U., Bareth, G., 2017. Multi-Data Approach for remote sensing-based regional crop rotation mapping: a case study for the Rur catchment, Germany. Int. J. Appl. Earth Obs. Geoinformation 61, 55–69. https://doi.org/10.1016/j.jag.2017.04.009.

Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. Times 222, 303–317. https://doi.org/10.1016/j.rse.2018.12.026.

Wang, S., Di Tommaso, S., Deines, J.M., Lobell, D.B., 2020. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. Sci. Data 7, 307. https://doi.org/10.1038/s41597-020-00646-4.

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. Remote Sens. Environ. Times 112, 1096–1116. https://doi.org/10.1016/j.rse.2007.07.019.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. Times 236 (111402). https://doi.org/10.1016/j.rse.2019.111402.

Wilson, B.T., Knight, J.F., McRoberts, R.E., 2018. Harmonic regression of Landsat time series for modeling attributes from national forest inventory data. ISPRS J. Photogrammetry Remote Sens. 137, 29–49. https://doi.org/10.1016/j.isprsjprs.2018.01.006.

World Food Programme, 2022. Burkina Faso Country Brief. Ouagadougou Burkin Faso: WFP. January: https://docs.wfp.org/api/documents/WFP-0000138348/download/?_ga=2.28720052.941926027.1649681483-292647639.1649681483. Available at.

Zafari, A., Zurita-Milla, R., Izquierdo-Verdiguier, E., 2017. Integrating support vector machines and random forests to classify crops in time series of Worldview-2 images. In: Bruzzone, L., Bovolo, F., Benediktsson, J.A. (Eds.), Image and Signal Processing for Remote Sensing XXIII. Presented at the Image and Signal Processing for Remote Sensing. SPIE, Warsaw, Poland, p. 34. https://doi.org/10.1117/12.2278421.

Zurita-Milla, R., Izquierdo-Verdiguier, E., de By, R.A., 2017. Identifying crops in smallholder farms using time series of WorldView-2 images, in: 2017 9th international workshop on the analysis of multitemporal remote sensing images (MultiTemp). In: Presented at the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, Brugge, Belgium, pp. 1–3. https://doi.org/10.1109/Multi-Temp.2017.8035246.